ESSAYS ON ECONOMIC BEHAVIOR OF LOAN REPAYMENT: EVIDENCE FROM HOUSEHOLD AND CORPORATE LOAN MARKETS

HE JIA

NATIONAL UNIVERSITY OF SINGAPORE

2013

ESSAYS ON ECONOMIC BEHAVIOR OF LOAN REPAYMENT: EVIDENCE FROM HOUSEHOLD AND CORPORATE LOAN MARKETS

HE JIA

(B.Sc. Eco. NKU)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF REAL ESTATE

NATIONAL UNIVERSITY OF SINGAPORE

2013

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety.
I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Hejia

He Jia

May 2014

Acknowledgements

I would like to express my deepest sense of gratitude to my supervisor, Professor Yongheng Deng, for his enormous support, guidance and encouragement throughout the course of my Ph.D study. I still remember the hours and hours of intellectual conversations and discussions with him. His words and guidance have always encouraged and inspired me to a higher level of thinking. This thesis would not have been possible without his intellectual inspiration, valuable comments, constant quality requirements and patient help. His rigorous attitude towards academic research and positive attitude towards life will benefit me all my life.

I also thank the committee members of my thesis: Associate Professor Sing Tien Foo and Associate Professor Xudong An. Professor An helped me tremendously on my thesis and research, especially for Chapter Four. He personified the spirit of a scholar to me during our joint project. Professor Sing gave me the opportunity to work together with him on several joint projects, and shared his profound knowledge with me. His patient help and encouragement have supported me in solve many problems.

My special thanks go to Associate Professor Wu Jing, who has discussed interesting research ideas with me, and worked together with me on several research projects. He has provided much help with my research, life and career development. He is a good mentor, and also a precious friend.

I would also thank Professor Sumit Agarwal. I have learnt a lot from our joint projects, and his acuminous mind and enthusiasm about academic work has touched me deeply.

I want to thank Associate Professor Fu Yuming, Professor Ong Seow Eng, Associate Professor Yu Shiming, Associate Professor Liao Wen-Chi, Dr Seah Kiat Ying, and all other faculty members who have helped me in Real Estate Department at National University of Singapore. A very special gratitude goes to Associate Professor Tu Yong, the Director of the Ph.D Programme in the Real Estate Department (NUS). Without her help and encouragement, I could not have started or finished my Ph.D studies. I have to thank Mr Quanlin Gu from Peking University for his help with Chapter Three.

I am very grateful for all helpful comments from the participants at the 2013 and 2014 American Real Estate and Urban Economics Society Annual Conference (AREUEA); the 2013 Asia Real Estate Society Annual Conference (AsRES); the 2012 and 2013 Global Chinese Real Estate Congress Annual Conference (GCREC); and the 2013 Asia Meeting of the Econometric Society.

I am also indebted to all my classmates and friends for their great help and invaluable discussions throughout my graduate studies. In particular, I would like to thank Wang Yourong, Zhou Xiaoxia, Guo Yan, Zhang Liang, Li Qing, Qiu Leiju, Zhang Bochao, Lai Xiongchuan. I also want to express my deepest thanks to a group of my friends in Singapore, including Li Hao, Dai Lingyun, Na Zhenkun, Yuan Xin, etc.

A doctoral life is often described as a solitary endeavor. However, my friends at the Institute of Real Estate Studies (IRES, NUS) made the Ph.D journey much more enjoyable. I would like to thank all members in IRES: Liu Xin, Luo Chenxi, Wang Yonglin, Zhao Daxuan, Li Congmiao, Liu Xueliang, Gu Fengjuan, Ren Chaoqun, Associate Professor Lum Sau Kim, Ms Ko Chen, Mr Kevin Chew Sung Yeow, Ms Zhou Xuefeng and Ms Nor Rahizah Binte Abdul Rahim.

I would like to thank the administrative staff members in Real Estate Department, especially Ms Zainab Bte Abdul Ghani and Ms Zheng Huiming, and Ms Nor'aini Bte Ali and Ms Wong Mei Yin in School of Design and Environment, who gave me a lot of help during my Ph.D studies.

This dissertation could not have been finished without the love, encouragement and support of my family. I would like to thank my parents, Bian Laiquan and An Lane, who have always supported and encouraged me silently, and my old brother, He Peng, who is always my idol and hero in my heart.

I want to thank my father-in-law, Tian Haiming, and mother-in-law, Wang Aiping for their unending encouragement and support.

Finally, I want to thank the person who supported me more than anyone else – my husband Tian Yu. Thank you so much for understanding and supporting me, and for always giving me strength and hope.

Table of Contents

Summary

This study enriches our understanding of three aspects of the economics of loan repayments by households and corporations: (1) residential mortgage default behavior, (2) residential mortgage partial prepayment behavior, and (3) corporate default risk.

The first essay studies whether heterogeneity in borrowers' time preferences correlates with their decision to default on their mortgage payments. Option theory predicts that borrowers should immediately exercise the default option when the market value of their mortgage exceeds the value of the underlying property. However, empirical evidence shows that a substantial number of borrowers are unlikely to default as 'ruthlessly' as option theory predicts. This indicates that mortgage borrowers are a heterogeneous group. In this essay, borrowers' time preferences across mortgage choices are hypothesized as being heterogeneous. Borrowers can either have a present-biased preference (overvaluing immediate outcomes), or a time-consistent preference (standard exponential discounting). Borrowers with a present-biased preference are more likely to accept back-loaded mortgages that minimize up-front costs, even though this increases their risk of going "underwater" and entering default when negative home price shocks occur.

The second essay studies the likelihood of making partial prepayments of mortgages and the process through which mortgage borrowers learn to make partial prepayment decisions in the residential mortgage market in China. The learning dynamics are measured by studying the repeated mortgage partial prepayment behavior of individual borrowers. As with full prepayments,

partial prepayment decisions impact the duration and pricing of mortgage-backed securities (MBS). However, unlike full prepayment, partial prepayment does not lead to a termination of the mortgage contract, allowing borrowers to repeat their actions in the future and learn from their early partial prepayment experiences. In the empirical tests, a longitudinal discrete choice model of the choice of mortgage payment is presented and estimated using a rich set of mortgage loan history data from a leading mortgage lender in China. The results indicate that path dependency and reinforcement learning arise whenever a borrower's partial prepayment decision depends not only upon current stage variables and his/her individual characteristics, but also on the learning experience (both from self and others). Borrowers with more partial prepayment experience in earlier stages have a higher probability of making the same decision in the future. Moreover, learning dynamics are not monotonic, and recent experience plays a larger role than distal experiences in determining a partial prepayment decision.

The third essay studies information beyond accounting and market variables in predicting the default of Chinese-listed companies. Although predicting the credit risk and the probability of default of companies is important, it is a challenging task for developing countries where information quality is poor. This essay comes up with a default probability prediction model for Chinese-listed companies that addresses information quality issues. The results indicate that, while accounting and market variables provide useful information about corporate default, other variables, such as state ownership, the real effective exchange rate, the money supply, the short-term lending rate, the coincident index and inflation, provide additional information. These variables increase the predictive power of accounting and market variables assessing corporate default. In addition, the default risk of Chinese companies are correlated and clustered within their industry. Certain industries, such as communications and technology, tend to have a higher default risk than others.

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Background

Loan repayment behavior and the riskiness of borrowers have received much attention in recent years. Borrowers can vary from corporations to households who borrow to finance their purchases of a home. Studying the behavior of debtors in repaying their financial obligations is important for lenders and investors.

A house is the single most important consumption good for a household, as well as being the dominant component of its wealth. However, since the value of a house usually exceeds a household's ability to pay, households generally pay for their homes through mortgage contracts, in which they borrow from mortgage lenders. During the payment process, mortgagors can choose to either default or fulfil prepayments to terminate their mortgage contracts. Prepayments may be full or partial. Prepayments in full mean that borrowers pay off the mortgage completely before it matures. Partial prepayments, where borrowers pay off part of their mortgages, have been less studied in the literature. Unlike default and full prepayment, partial prepayments do not terminate mortgage contracts, but instead change the unpaid mortgage balance. [1] Borrowers can carry out partial prepayments repeatedly. Understanding the different ways that mortgage borrowers decide to repay their mortgages is crucial for both mortgage lenders and investors of mortgage-backed securities.

Most research on corporate borrowing has focused on corporate default risk. The credit risk of corporations is the oldest form of risk in financial markets, and it still continues to attract fervent interest among academics, practitioners and regulators. A typical method of corporate borrowing is corporate bonds, which are debt instruments issued by corporations. When they issue bonds, corporations promise to make specified payments to their bondholders based on the dates and amounts fixed in a contract. However, bondholders may not receive the promised payments in full or in part, because corporations may default on their agreement, causing bondholders to experience financial loss. Effectively estimating corporate default is crucial to those responsible for granting bank loans or investing in financial products exposed to corporate default risk.

1.1.1 Loan Repayment Behavior of Household

Residential mortgages are a debt instrument in which an individual borrower uses real property as collateral to obtain a lump sum of money from a mortgage lender to finance a property purchase. In this process, according to the contract between the borrower and the lender, the borrower should make

[1] Partial Prepayment is defined as any extra money in addition to the monthly mortgage payment, and the extra amount paid will be used toward the reduction of the unpaid balance of the loan. In China, by making mortgage partial prepayment, borrowers can choose to shorten the horizon of their remaining loan term without lessening the regular monthly payments, or they can choose to reduce their monthly payment instead of shortening the remaining payment term.

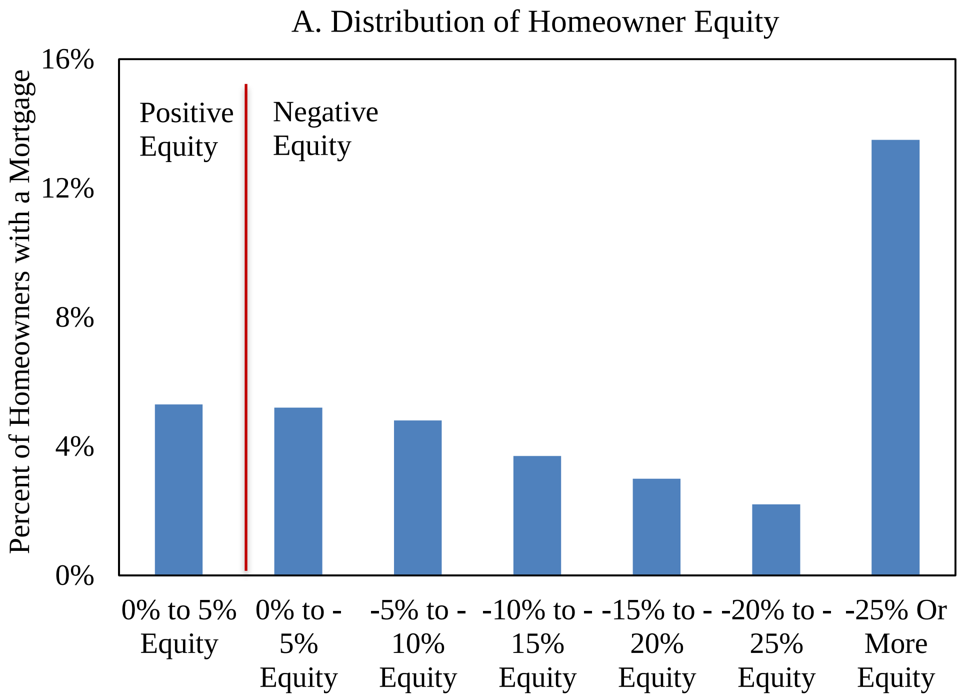
monthly payments to the lender based on the mortgage contract. A mortgage can enter default if the borrower does not make the monthly payment. In this case, the lender can choose to take over the property to recover the balance of the loan.

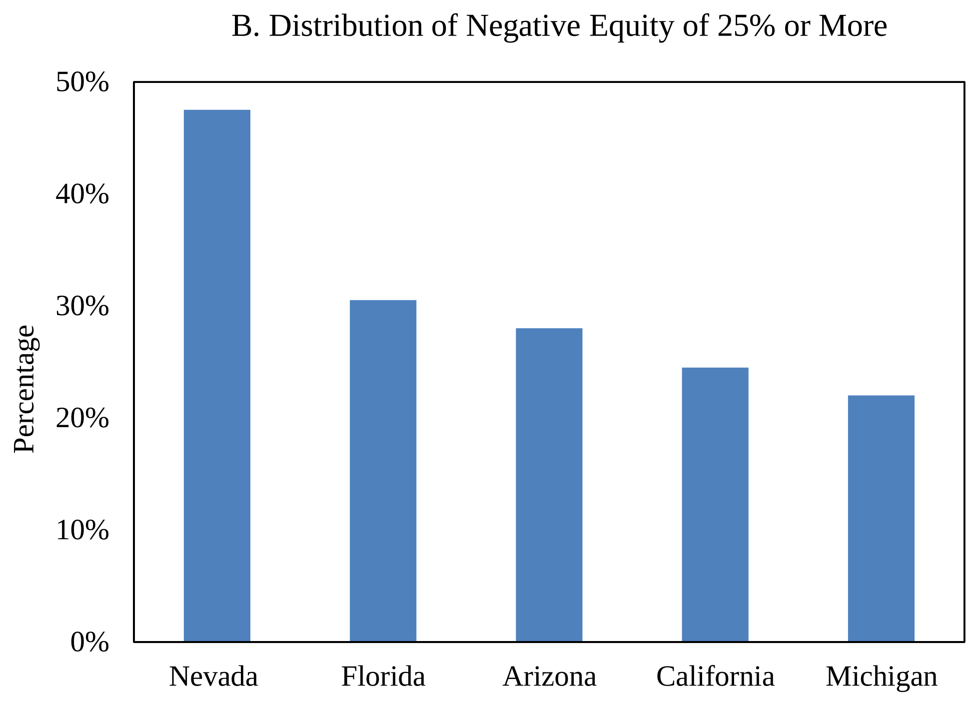
Defaulting on a mortgage is associated with substantial costs. First, mortgage default is costly to borrowers. By defaulting on their mortgages, borrowers will be penalized with a lower credit score and fewer opportunities to apply for mortgages and purchase homes in the future, as well as the emotional distress and moving costs. On the other side, mortgage default is also costly for mortgage lenders and investors. The net value of compensation from foreclosure is generally less than the asset value, and lenders will suffer a financial loss from a borrower’s decision to default on their mortgage and the ensuing foreclosure process (Quercia and Stegman 1992). Investors holding mortgage-backed securities are also exposed to the default risk of the mortgagor. After the financial crisis began in the last quarter of 2007 with the collapse of housing prices in the United States and the widespread rise in the default rates of both prime and subprime mortgages, more and more attention has been paid to mortgage default and foreclosures by scholars, policy analysts, and government officials.

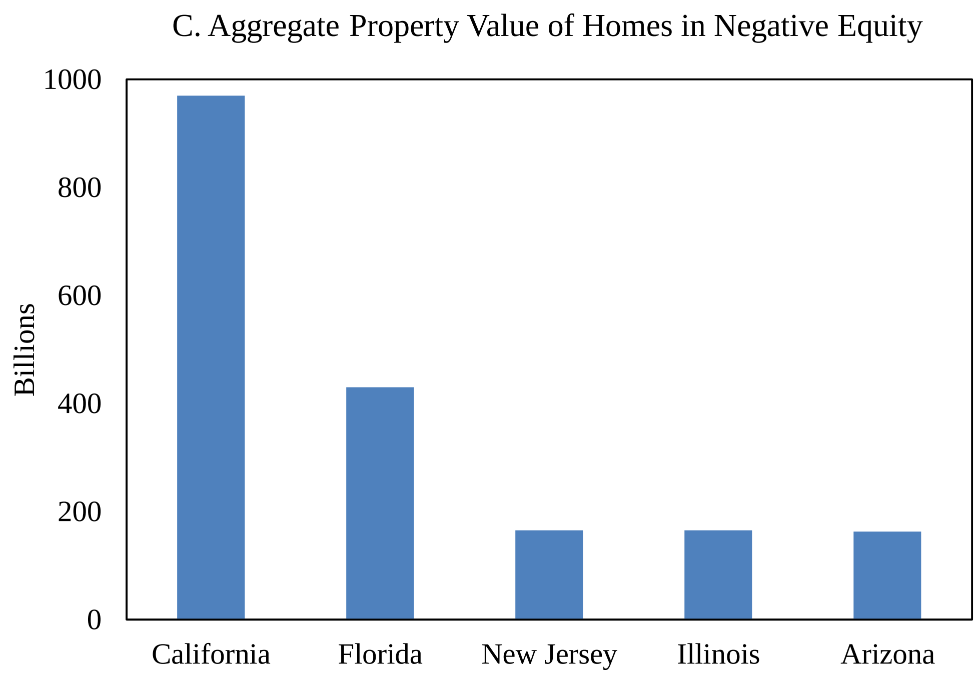
Under the “ruthless” hypothesis proposed by the contingent claims model, the well-informed borrower will default immediately when the mortgage value exceeds the property value at any time during the loan term (Titman and Torous 1989; Kau et al. 1992). Thus, the negative equity position of the homeowner is a key factor in affecting a borrower’s decision to default. Figure 1-1 shows the distribution of homeowner equity in the second quarter of 2009.

There were nearly 32.2% of all mortgage properties in negative equity as of June 30, 2009 in US. Nevada had the highest percentage with nearly two thirds of mortgage borrowers in a negative equity position. For Arizona, Florida, Michigan, and California, the percentage of negative equity were 51%, 49%, 48%, and 42% respectively. The total negative equity shares of these top five states were nearly 47% of the total negative equity in US. The more negative equity is, the higher the risk that the homeowners will lose their home. The scourge of negative equity and its quantitatively important role on default cannot be neglected.

Figure 1-1 Distribution of Homeowner Equity in US







Note: From Negative Equity Summary Report of CoreLogic, 2009 quarter two. Panel A is the equity distribution among the whole nation. Panel B is the distribution of negative equity of 25% or more among the top five states in US: Nevada, Florida, Arizona, California and Michigan. Panel C is the distribution of aggregate property value of homes in negative equity of the top five states in US: California, Florida, New Jersey, Illinois and Arizona.

Numerous papers have tried to address the triggers that contributed to the increasing number of mortgage defaults and foreclosures in the recent financial crisis. Some have pointed to lax underwriting standards, arguing that mortgage securitization reduced the lender's incentive to be prudent, leading to lower underwriting standards in mortgage origination and eventually causing lower loan quality and higher default rates (Keys et al. 2010; Agarwal, Chang, and Yavas 2012; Keys, Seru, and Vig 2012). Still others have argued that irrational expectations regarding future house price growth should also be responsible for contributing to this subprime crisis. For instance, Mian and Sufi (2009) pointed out that lenders' increased expectations of future house price growth may have been responsible for the increase in subprime mortgage credit. Other explanations include the rise of subprime lending, decline of risk premiums, and problems with credit rating agencies.

However, compared to the number of borrowers who simply 'walked away' from their homes and those who defaulted, a large proportion of borrowers with negative equity continued to make their mortgage payments, even their mortgages were deeply underwater, including those who lived in "nonrecourse states" such as California and Arizona.[2] The fact that many borrowers continued repaying their underwater mortgages so as to keep their homes challenges traditional option models of well-informed borrowers operating in a world without economic frictions (see Vandell 1995). In a similar vein, Quigley and van Order (1995) found that there were no costs to default, other than the loss of the house in the "ruthless" default model. They emphasized

[2]In non-recourse states, lenders cannot pursue defaulting homeowners for a deficiency judgment. While lenders can recover some of their losses by foreclosing on the property, they cannot sue borrowers for additional funds. If the foreclosure sale does not generate enough money to satisfy the loan, lenders must accept the losses.

the existence of transaction costs and their impact on default decisions. More recently, White (2010b) hypothesized that the shame and guilt of foreclosure and the large perceived penalties for defaulting keeps borrowers from default when it would be in their financial interest to exercise the option. Indeed, Guiso, Sapienza, and Zingales (2013) found that mortgage borrowers tended to view default as immoral, and continued to make their mortgage payments even when their mortgages were deeply underwater.

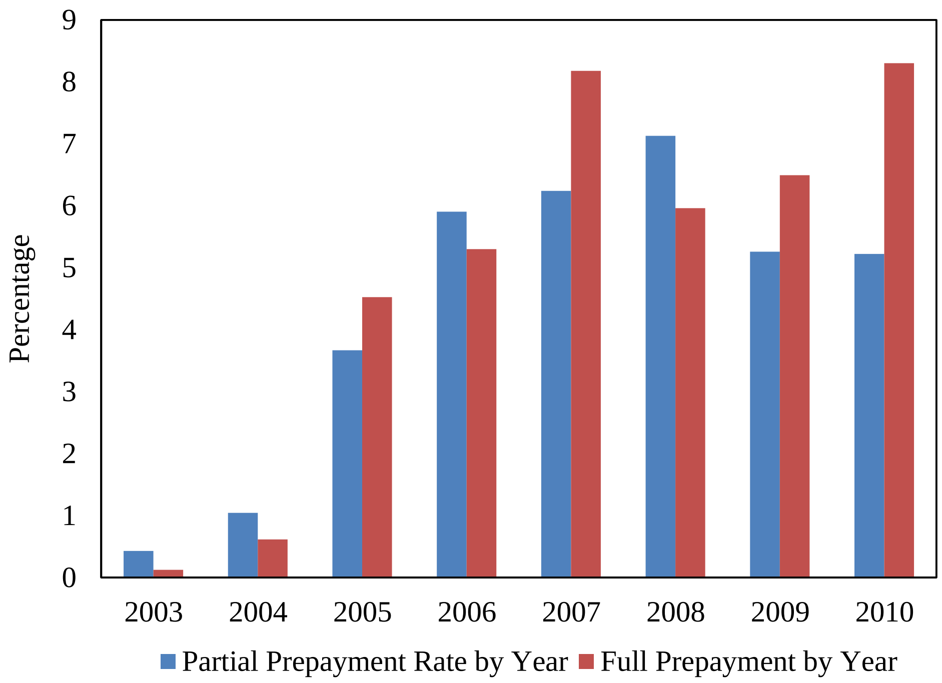
Why would some homeowners choose to default on their mortgages while others continued to make payments on houses worth less than their mortgages? Chapter Two of this study focuses on this heterogeneous mortgage default behavior, as well as the underlying theory of unobserved heterogeneity among mortgage borrowers and the source of this unobserved heterogeneity.

Besides choosing mortgage default, borrowers can also fully pay off their mortgages to terminate the mortgage contract. Full prepayment risk and default risk are the two most important types of termination risks, and have been examined in many studies, along with the resultant behavior of borrowers (Kau et al. 1992; Stanton 1995; Deng, Quigley, and Van Order 2000). A number of studies have explained either default or prepayment behaviors (Dunn and McConnell 1981a, 1981b; Quigley and Van Order 1991; Lekkas, Quigley, and Van Order 1993), while some papers have addressed them jointly (see Deng et al. 2000).

Compared to default and full prepayment behaviors in the residential mortgage market, relatively few papers have examined mortgage partial prepayment behavior. A partial mortgage prepayment occurs when a borrower repays some, but not all, of a housing loan ahead of schedule. Just like full prepayment risk, partial prepayment also introduces risk to the duration of mortgage-backed securities (MBS), and affects their pricing. The uncertainty of a borrower's partial prepayment decision increases the risk to mortgage lenders. In addition, investors holding portfolios of mortgage-backed securities are also exposed to the partial prepayment risk. However, a borrower's partial prepayment decision is different from full prepayment, as it only changes the unpaid mortgage balance but does not terminate the mortgage contract. This allows the borrower to repeatedly engage in this behavior and learn from his/her early partial prepayment experience when making future decisions.

Although partial prepayments may not be popular in the West, they can be frequently found in some regions and countries, such as China. Due to the absence of prepayment penalties, mortgage prepayment behavior (both partial prepayment and full prepayment) is very common in China (Figure 1-2). Partial prepayment of a mortgage introduces risk to mortgage securitization, with unique characteristics that are different from full prepayment behavior. As the mortgage market in China matures and becomes more competitive, there is an increasing need to understand the behavior of partial prepayment. Chapter Three of this study focuses on risk of partial prepayment among mortgage borrowers, and the processes through which mortgage borrowers learn to make partial prepayment decisions.

Figure 1-2 Partial Prepayment and Full Prepayment Rates in China Mortgage Market by Year



Note: From a leading mortgage lender in China. Y-axis measures the full prepayment rate and partial prepayment rate, while X-axis measures the year.

1.1.2 Corporate Loan Repayment Behavior

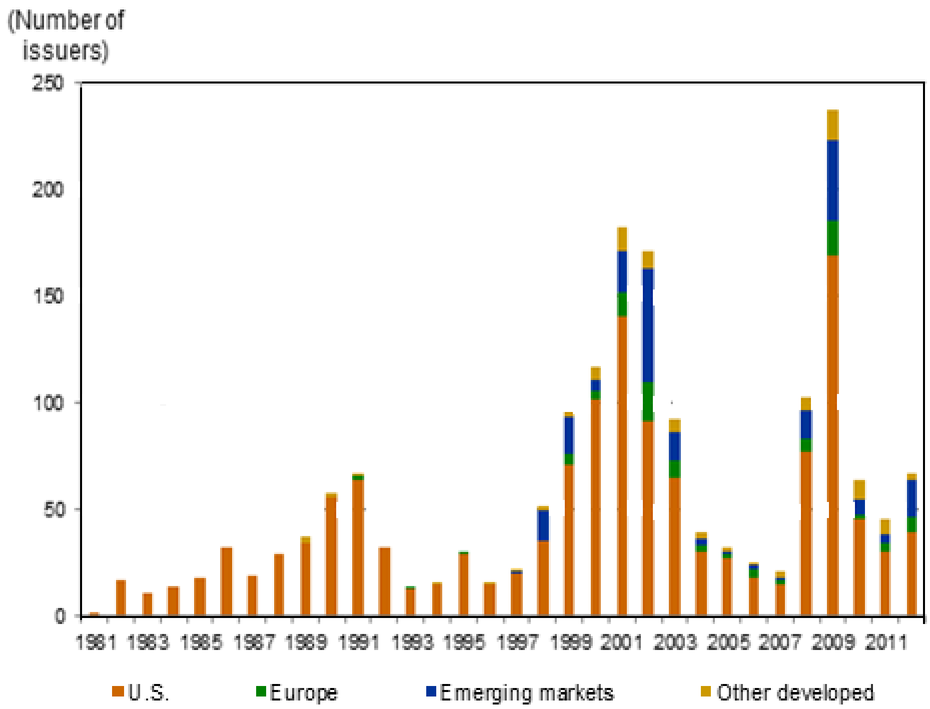
The recent global financial crisis and the increased number of corporate defaults emphasize the importance of corporate credit risk management. A corporation is in default when it fails to make specified payments obligated to its creditors. Examples of corporate default include a bond default and corporate bankruptcy. Corporate defaults lead to financial losses for a corporation's creditors, and also have a negative impact on society and the overall economy. Understanding which variables are relevant for predicting corporate default risk is an important issue for academics, practitioners and regulators.

Researchers have used a variety of variables to predict corporate default in reduced-form corporate bankruptcy forecast models, which have proven to be a useful tool for predicting corporate bankruptcy (Altman 1968; Ohlson 1980; Zmijewski 1984; Beaver, McNichols, and Rhie 2005). These models have two main objectives: identifying the relevant predictive variables and improving prediction accuracy.

Earlier studies, such as Beaver (1966), Altman (1968), Ohlson (1980), and Zmijewski (1984), used accounting ratios as a measure of default risk. These ratios are derived from a firm's operating environment and include size, book-to-market equity ratio, and growth rate. More recently, researchers have proposed to use market-driven variables, derived from a firm's information and trading environments, instead of accounting ratios to predict corporate bankruptcy. Their choice is based on the argument that a firm's default probability should be perfectly reflected at all times in the market value of its equity (Shumway 2001; Chava and Jarrow 2004; Hillegeist et al. 2004; Campbell, Hilscher, and Szilagyi 2008).

Emerging markets are becoming increasingly important engines of global economic growth. As their economies have developed, their corporate default rates have stabilized in recent years (Figure 1-3), especially after the Asian financial crisis in 1997-8. However, within developing countries, information quality becomes a potential issue when applying both the accounting-based model and the market-driven model to predict corporate default. Developing countries, including those in Asia, are prone to poor accounting and legal standards. Moreover, institutional effects play a unique role in several Asian markets, making stock prices less informative in these markets. As one of the leading emerging economies, China is witness to all of these concerns.

Figure 1-3 Annual Corporate Defaults by Number of Issuers



Note: From Standard & Poor's Global Fixed Income Research and Standard & Poor's CreditPro. Count excludes defaults that were not rated one year prior to default. Other developed includes Australia, Canada, Japan, and New Zealand.

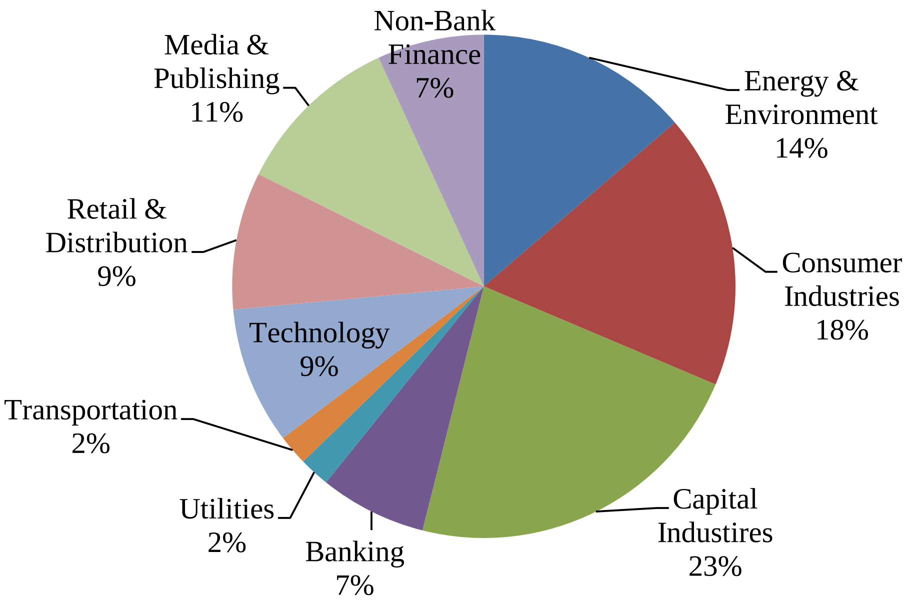
In addition, more and more attention has been paid to the performance of the Chinese state-owned enterprises (SOEs), which play a very important and unique role in China's economy. Since the founding of the People's Republic of China, SOEs have undergone several waves of reform. Although the number of SOEs has decreased, the important role of SOEs in Chinese economy has not changed (Deng et al. 2011). Except for their rentability, which is similar to that of the non-SOEs, SOEs also have a special social function in China. For example, they should serve the macro-economy to optimize the allocation of resources within the scope of the whole society, balance regional economic development to realize the rational layout of economy, and control the lifeline and other important areas of the national economy to protect China's economic, political and military security. The special role of state ownership may lead to corporate default in China having some unique characteristics.

According to Moody's (2010), significant industry fixed-effects exist in the bankruptcy of US firms (Figure 1-4). It can be seen from Panel A, that in 2010, almost 23% of the defaults occurred in the debt issued by the Capital Industries sector.[3] Issuers in the Consumer Industries sector accounted for the next highest share of defaults at 18%.[4] Although the Capital Industries sector accounted for 23% of defaults in 2010, it was not the sector with the highest rate of default. That distinction belongs to the Media & Publishing industry, which had a 3.6% default rate in 2010. Chava and Jarrow (2004) emphasized that industry groupings have significant effects on the forecasting equations of corporate default. The similar industry fixed-effects are supposed to be existed for Chinese listed companies.

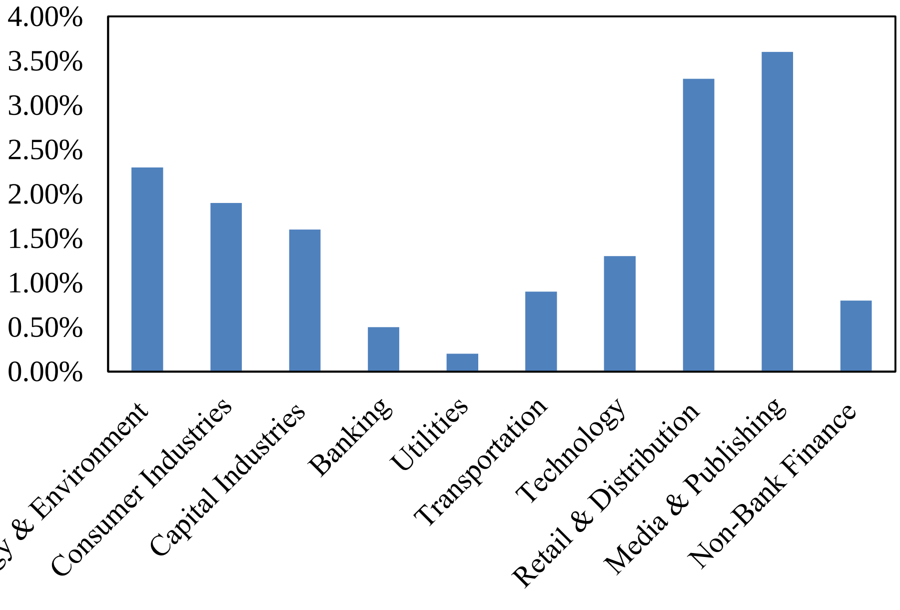
[3] Capital industries sector includes automotive, capital equipment, chemicals, plastics & rubber, construction & building, containers, packaging, & glass, forest products & paper, metals & mining, and business service industries.
[4] Consumer industries sector includes beverage, food, & tobacco, durable & non-durable consumer goods, healthcare & pharmaceuticals, hotel, gaming, & leisure, and consumer service industries.

Figure 1-4 Defaults by Broad Industry in US

Panel A Percent of Issuer Counts



Panel B Default Rate by Broad Industry



Note: From Moody's 2010.

Overall, information beyond accounting and market variables needs to be explored in predicting the default probabilities of Chinese-listed companies.

Chapter Four of this study focuses on the information required to predict the corporate default of Chinese listed companies.

1.2 Research Objectives

This research aims to explore three main aspects of the loan repayment behavior of households and corporations: residential mortgage defaults, partial prepayments of residential mortgages, and corporate default risk.

The first essay examines whether heterogeneity in borrowers' time preferences is correlated with their decision to default on their mortgage by measuring borrowers' time preferences using their mortgage type choice. In particular, it investigates whether borrowers who exhibit present-biased preference – those who are more likely to accept back-loaded mortgages that minimize up-front costs – have a higher chance of their mortgages going "underwater" and into default following negative home price shocks. The research questions for the first essay are *"What are the origins of borrowers' unobserved heterogeneity?"* and *"Do heterogeneous borrowers' time preferences correlate with their default decisions?"*

The second essay explores the risk of partial prepayment and the reinforced learning process of borrowers in the Chinese mortgage market. It aims to add to the understanding of the partial prepayment behavior of borrowers, which has been previously neglected in the literature. This essay also aims to provide strong evidence to support the important role of learning in explaining partial prepayment behavior using individual loan-level data. Lastly, since emerging markets are fast becoming important engines of the global economy, it is important for more research on them. This study contributes to this by examining the mortgage market in one of the most important emerging markets, China. The research questions for the second essay are *"What factors affect the partial prepayment decisions of Chinese mortgage borrowers?"* and *"Is there any learning process in the partial prepayment decisions?"*

The third objective of this study is to build a bankruptcy prediction model that overcomes the data limitations of Asian companies, especially Chinese-listed companies. This model will make use of information beyond accounting and market variables to predict the default probability of Chinese-listed companies. Specifically speaking, there are four main research questions. First, *"How useful are accounting and market variables in predicting the default probability of Chinese-listed companies?"* Second, *"What is the impact of state ownership on corporate default after controlling for variations in the financial performances of observable firms?"* Third, *"Do macroeconomic variables provide additional information that is beyond firm-specific accounting and market information on corporate default?"* Finally, *"How is the default risk of Chinese-listed companies correlated and clustered?"*

1.3 Knowledge Gaps

1.3.1 Time Preferences, Mortgage Choice and Mortgage Default

Over the past thirty years, substantial studies have been proposed to explain default risk and the default behavior of mortgagors. Research in this area is mainly based on option theories, which provide useful frameworks for analyzing borrowers' behavior. The "pure" option-based theoretical mortgage pricing models assume that well-informed borrowers will default 'ruthlessly' once their mortgage value exceeds their property's value at any time during the term of their loan (Titman and Torous 1989; Kau, Keenan, and Kim 1993). Nevertheless, while negative equity may be a necessary condition to trigger default, it is not a sufficient one (Vandell 1995; Deng et al. 2000; Bajari, Chu, and Park 2008; Bhutta, Dokko, and Shan 2010). Exercising the default option means to give up the option to default and refinance later on (Kau, Keenan, and Kim 1994). Empirical evidence shows that substantial borrowers are less likely to default as 'ruthlessly' as the option theory predicts.

These pure option-based models build on perfectly competitive markets without any transaction or reputation costs, and with no exogenous reasons for residential mobility. However, the idea of a "frictionless" market in option theory is an ideal assumption, since transaction costs have a significant impact and are expected to cause individuals to strategically default or prepay. Substantial empirical research using the option-based framework has attempted to illustrate and explain this phenomenon by arguing that the transaction costs resulting from default are pervasive and significant for borrowers (Stanton 1995; Archer, Ling, and McGill 1996; Harding 1997). The literature assumes that the transaction costs, such as moving costs, reputational issues, and default penalties, are sufficiently high to force homeowners to leave their homes. In addition to the economic considerations of transaction costs, attention has also been focused on the emotional constraints to strategic default. Using survey data, Guiso et al. (2013) documented that social and moral considerations may partially play a role in explaining the willingness of homeowners to continue paying their underwater mortgages.

The role of transaction costs is important in determining the exercise of both default and prepayment options. What causes these strategic default borrowers to accept the economic and emotional transaction costs? Empirically, the unobserved heterogeneity of mortgage borrowers is extensively discussed in existing literature (Stanton 1995, 1996; Deng et al. 2000; Hall 2000).

Given the above, transaction costs and unobserved heterogeneity have a clear role when examining mortgage default. However, unobserved heterogeneity is still a black box that attracts much attention. Researchers have been unable to provide an underlying theory to explain the unobserved heterogeneity of borrowers or its origins. Chapter Two in this study tries to fills this gap by assuming that borrowers' time preferences are heterogeneous, and shows that heterogeneity in borrowers' time preferences correlates with their mortgage default decisions.

1.3.2 Reinforcement Learning and Mortgage Partial Prepayment Behavior

In the mortgage market, prepayment risk and default risk are the most important two general types of borrower's termination risk. The literature has widely studied these risks and the corresponding borrower's behavior (Kau et al. 1992; Stanton 1995; Deng et al. 2000). Most studies are based on the contingent claims models, where prepayments are treated as American call options and defaults as compound put options. While some studies pay attention only to full prepayment behavior (Dunn and McConnell 1981a, 1981b; Buser and Hendershott 1984; Brennan and Schwartz 1985; Schwartz and Torous 1989), others focus only on mortgage default risk (Cunningham and Hendershott 1984; Foster and Van Order 1984, 1985; Vandell and Thibodeau 1985; Titman and Torous 1989; Quigley and Van Order 1991; Kau, Keenan, and Kim 1993; Lekkas, Quigley, and Van Order 1993). A series of papers has emphasized the importance of examining full prepayment and default options jointly, and treated them as competing risks in determining mortgage termination (Kau et al. 1994; Kau and Keenan 1996; Deng, Quigley, and Van Order 1996; Deng 1997; Deng et al. 2000).

Relative to the extensive literature on default and full prepayment behavior in the residential mortgage market, only a few papers have been conducted on the partial prepayment behavior of borrowers. Hayre and Lauterbach (1991) were the first to emphasize the distinct role of partial prepayment behavior, which is unlike full prepayment, and they tried to capture the effect of partial prepayment by adding an average constant dollar amount each month in the prepayment model. Chinloy (1993) built both theoretical and empirical models which treat full prepayment and partial prepayment as separate decisions to dislodge the analysis bias. Abrahams (1997) discussed the effect of partial prepayment on full prepayment modelling. Overall, only a few papers pay attention to partial prepayment, relatively to default and full prepayment behavior in the mortgage literature.

Learning is a multi-disciplinary area which draws much attention from economics, psychology, cognitive science, computer science, mathematics, and neural science (Bush and Mosteller 1955; Cross 1973; Arthur 1991, 1993; Roth and Erev 1995; Erev and Roth 1998). There has been growing interest in the effects of learning and experience on an individual decision-making. However, because of data limitation, most economic studies have analyzed learning either in laboratories or in the field. Only a few papers have examined learning using individual micro-level data.

Since 1998, the Chinese housing market has developed rapidly and undergone significant reform. Correspondingly, the Chinese residential mortgage market has grown quickly and continues to attract attention. However, little research has been conducted on the Chinese mortgage market because of data limitations.

In short, in the literature on mortgage payment behavior, there exists a gap in our understanding of the partial prepayment risk of borrowers and their associated behavior. In the literature on learning, there is a gap in studying learning using micro-level data. Chapter Three in this study fills these two gaps by modelling the learning behavior of borrowers through repeated mortgage partial prepayment decisions. In addition, Chapter Three also contributes to the literature on the Chinese mortgage market.

1.3.3 Information beyond Accounting and Market Variables in Predicting Corporate Default

Accurately predicting the probability of corporate default is of great interest to all academics, practitioners and regulators. Research in this area mainly focuses on Western markets, and studies vie to increase the accuracy of their default predictions by using different mixes of accounting-based and market-driven variables.

Beaver’s (1966) pioneering work used a business failure prediction model based on financial ratios derived from corporate financial reports to show that certain financial ratios were statistically significantly related to corporate failure. Building on that work, a huge body of literature has been generated to predict corporate default using different accounting-based variables. For example, Altman (1968) extended Beaver's (1966) work by investigating a set of financial and economic ratios into a Multivariate Discriminant Analysis (MDA) model to derive the Z-score measure for predicting bankruptcy. Taffler (1983) used the same technique to generate the UK-based Z-score. Altman, Haldeman, and Naraganan (1977) used Quadratic Discriminant Analysis (QDA) which incorporated comprehensive inputs to identify bankruptcy risk of firms. Ohlson (1980) applied a conditional logit model to predict corporate default (known as "O-score"), which included seven accounting-based explanatory variables. Zmijewski (1984) performed a probit model based on accounting data but uses a different set of independent variables as previous studies. Lau (1987) constructed a five-state financial distress prediction model, and recognized a firm would enter each of the five states of financial distress using a multinomial logit model. However, models based on accounting variables are criticized because they are backward-looking, prone to manipulation, and underestimate book value.

Compared to backward-looking accounting ratios, the information in stock prices tends to be more forward-looking. Therefore, market-driven variables have been used in default prediction models to improve the accuracy of their forecasts. Black and Scholes (1973) and Merton (1974), used an option-pricing method, and initiated the large amount of research on the use of market-based variables. Key studies in this area include Crosbie and Bohn 2002), Hillegeist et al. (2004), Vassalou and Xing (2004), Reisz and Perlich (2007), and Bharath and Shumway (2008).

However, both accounting-based models and market-driven models cannot be used in developing countries because of their poor accounting and legal standards and less informative stock prices (Shleifer and Robert 1994; Morck, Yeung, and Yu 2000). These concerns about information quality extend equally to China, one of the leading economies in the emerging market.

Given the above mentioned concerns, although accounting ratios and market-driven variable discriminant analysis can be used to improve our understanding in Western markets, the same cannot be used in China. Information beyond accounting ratios and market variables is needed to predict the probability of Chinese companies entering default. Chapter Four of this study focuses on filling this gap.

1.4 Significance of the Study

In the first essay, setting borrowers' heterogeneous time preferences against their mortgage choices may improve our understanding of the variety of mortgage default behavior, and will help develop better policies to deal with the issues ensuing from the housing foreclosure crisis, such as mortgage modifications and mortgage contract design.

The second essay in this study will enhance our understanding of partial prepayment behavior, which has been neglected in the literature. In addition, the findings will provide valuable insights about the housing and mortgage markets in China, as well as those in other transition economies. Lastly, by studying the learning process in partial prepayment decisions, this essay will provide strong evidence to support the important role of learning in explaining partial prepayment behavior in the mortgage market.

The model that is developed in the third essay will be a useful tool for risk assessment and management. This essay will hopefully remind lenders and investors to be aware of other factors that affect lending and investment risks, and suggest that policy makers should take into consideration the impact of monetary policy on the likelihood of corporate default.

1.5 Structure of the Thesis

This thesis is organized into five chapters. Chapter One presents the research background, research objectives, research questions, and significance of the research. Chapter Two comprises the first essay, entitled "*Time Preferences, Mortgage Choice and Mortgage Default*", which examines the correlation between the heterogeneity in borrowers' time preferences and their default decisions. Chapter Three consists of the second essay, entitled "*Reinforcement Learning and Mortgage Partial Prepayment Behavior*". It explores the risk of residential mortgage partial prepayments and how learning reinforces the behavior of borrowers in residential mortgage markets. Chapter Four presents the third essay, entitled "*Predicting Default of Chinese Companies: Information beyond Accounting and Market Variables*". This chapter investigates the role of information besides accounting and market variables in predicting the default of Chinese-listed companies. The final chapter concludes this thesis, and summarizes its contributions. It also highlights the limitations of the study, and provides recommendations for future research.

Chapter 2 Time Preferences, Mortgage Choice and Mortgage Default

2.1 Introduction

The financial crisis of 2008 triggered by the stunning rise of subprime mortgage delinquencies has led to a re-evaluation of mortgage defaults and foreclosures. In the past five years, millions of homeowners in the United States "walked away" and allowed their homes to be foreclosed. Moreover, according to CoreLogic's negative equity report, owners of 11.1 million residential properties were in negative equity (i.e. they were "underwater") and at the risk of foreclosure by the end of 2011.[5] Table 2-1 lists the ten states with the highest levels of negative equity and near negative equity in the United States.[6] Compared with borrowers who simply "walked away" and those in default, most underwater homeowners still continued to make their mortgage payments. However, it is difficult to know beforehand which borrowers will be in default, because there is significant heterogeneity among them. This leads us to wonder why some homeowners choose to default on their mortgages, while others do not.

---

[5] http://www.corelogic.com/about-us/news/corelogic-reports-negative-equity-increase-in-q4-2011.aspx Accessed on March 15, 2012.

[6] Properties that are in negative equity, where the borrower owes more on their mortgage than the property's current market value, are often termed as being "underwater" or "upside down". Mortgages that are within five percent of being in a negative equity position are defined by CoreLogic as being "near negative equity".

Table 2–1 Negative Equity in Selected States of US

State	Negative Equity Share	Near Negative Equity Share
Nevada	56.90%	5.30%
Florida	42.10%	4.10%
Arizona	38.60%	5.10%
Georgia	35.60%	6.30%
Michigan	32.00%	4.80%
California	28.30%	4.50%
Illinois	25.40%	4.60%
Ohio	23.80%	5.70%
Maryland	22.90%	4.80%
Idaho	22.30%	5.30%

Note: From Negative Equity Summary Report of CoreLogic, Jan-2013.

This chapter studies whether heterogeneity in borrowers' time preferences correlates with their decision to default on their mortgage payments. Borrowers' time preferences are measured using their choice of mortgage type. In particular, it investigates borrowers who exhibit present-biased preference, that is, those who are more likely to accept back-loaded mortgages that minimize up-front costs and place them at greater risk of becoming underwater and default following negative home price shocks.

After the pioneering work by Asay (1978), there was a quick expansion of studies on mortgage valuation and borrower behavior based on the contingent claims models, mainly developed by Black and Scholes (1973), Merton (1973a), Cox, Ingersoll, and Ross (1985). The contingent claims model provides a useful framework for analysing borrowers' behavior, in which prepayment is treated as an American call option and default as a compound put option. The "pure" option-theoretic mortgage pricing models assume that a well-informed borrower will default immediately when the mortgage value exceeds the value of his/her property at any time during the loan term (Titman and Torous 1989; Kau, Keenan, and Kim 1994). These models assume a perfectly competitive market without any transaction or reputation costs, and no exogenous reasons for residential mobility. Despite a frictionless market being an ideal case, negative equity may be a necessary, but not a sufficient, condition to trigger default (Vandell 1995; Deng, Quigley, and Van Order 2000; Bajar, Chu, and Park 2008). Evidence shows that a substantial number of borrowers are unlikely to default as 'ruthlessly' as the option theory predicts. White (2010b) argued that not all homeowners who were underwater on their mortgage walked away from their home immediately during the recent financial crisis, including those who lived in "nonrecourse states", such as California and Arizona. Although such behavior may appear irrational on its face, the homeowners who stayed and those who walked away all struggled with the same decision: to continue paying their mortgage or not.

Many empirical studies have tried to explain this 'irrational' phenomenon using the option-based framework. The anecdotes underlying these studies emphasize that the transaction costs resulting from default are pervasive and significant (Stanton 1995; Archer, Ling, and McGill 1996; Harding 1997).[7] These studies assume that these transaction costs, which include moving costs, reputational issues, and default penalties, are high enough to deter homeowners from leaving their homes. In addition to the economic consideration of transaction costs, recent attention has been paid on the emotional constraints to strategic default. Guiso, Sapienza, and Zingales

[7]For an explicit discussion about these transaction costs, see Kau, Keenan, and Kim (1993).

(2013), using survey data, documented that social and moral considerations may play a partial role in explaining the willingness of homeowners to continue to pay their underwater mortgages. White (2010b) also argued that the shame or guilt associated with foreclosure and fear over the perceived consequences of foreclosure led those underwater homeowners choose not to default.

The role of transaction costs is important in determining the exercise of both the default and prepayment options. What causes default borrowers to accept the economic and emotional transaction costs that accompany their default decision? The empirical unobserved heterogeneity of mortgage borrowers has been discussed extensively in the literature, but only as an unproven assumption. For example, Deng et al. (2000) assumed that borrowers are heterogeneous agents who form discrete groups, and Hall (2000) assumed that these agents have different distributions of underlying hazards. Stanton (1995, 1996) and others have also argued that heterogeneity exists within mortgage pools. However, researchers have been unable to provide a theoretical framework for this unobserved heterogeneity or explain its origins.

The present study tries to fills this gap by assuming that this unobserved heterogeneity is based on borrowers' time preferences, and it examines whether heterogeneity in borrowers' time preferences is correlated with their mortgage default decisions. Two kinds of time preferences exist for borrowers, with corresponding discounting factors: present-biased preference (overvaluing immediate outcomes), and time-consistent preference (standard exponential discounting). The key distinction between these two is the presence and absence of a "present bias". Individuals with present-biased preference prefer immediate gratification, and, as a result, are more likely to minimize their up-front costs and postpone their mortgage payment. They are thus more likely to select back-loaded mortgages, such as interest-only loans. This selection is more likely to place them at a higher risk of going underwater and defaulting following negative home price shocks.

The hypothesis is made that naïve borrowers with present-biased preference are more likely to select interest-only loans which allow them to enjoy the immediate benefits of homeownership and postpone their mortgage payment costs. Sophisticated borrowers with present-biased preference, on the other hand, are fully aware of their future self-control problems, and know their future preference exactly, even though they may differ from their current preference. Therefore, they are smart and more likely to choose 30-year adjustable-rate loans. In contrast, borrowers with time-consistent preference tend to choose 30-year fixed-rate loans, which are fully amortizing mortgage loans where the interest rate on the note remains the same through the term of the loan.

Two main databases are used in this chapter: mortgage data at the level of individual loans, principally collected by BlackBox Logic (BBL), and home loan application and origination data, collected by the Home Mortgage Disclosure Act (HMDA). This chapter uses a logistic regression to examine how borrowers' time preferences influence their mortgage choice and their default decisions. The fixed effects are the year of origination, year of termination, and the state in which the property is located. Firstly, the default behavior of naïve borrowers who selected interest-only loans, relative to those who selected 30-year fixed-rate loans is studied. The results indicate that borrowers with 5-year interest-only loans are around 41 percentage points more likely to default than those who selected 30-year fixed-rate loans. In addition, borrowers with 10-year interest-only loans are around 47 percentage points more likely to default than dynamically-consistent borrowers who chose 30-year fixed-rate loans.

The further regression is done by studying the default behavior of borrowers who chose 30-year adjustable-rate loans relative to those who chose 30-year fixed-rate loans. Results indicate that sophisticated borrowers who chose 30-year adjustable-rate loans are 27 percentage points more likely to default than borrowers who selected 30-year fixed-rate loans. In other words, the default rate of sophisticated borrowers with present-biased preference is higher than borrowers with time-consistence preference.

Lastly, the default behavior of both interest-only loans and 30-year adjustable-rate loans relative to those who selected 30-year fixed-rate loans is examined. The results indicate that present bias is highly correlated with mortgage default, and borrowers who exhibit present-biased preference in their choice of mortgages have a substantially higher probability of default. Borrowers with 5-year interest-only loans are around 35 percentage points more likely to default than those who selected 30-year fixed-rate loans, and borrowers with 10-year interest-only loans are around 39 percentage points more likely to default than dynamically-consistent borrowers who chose 30-year fixed-rate loans. In addition, the default probability of borrowers with 30-year adjustable-rate loans is 24% higher than borrowers who chose 30-year fixed-rate loans. The association between present bias and mortgage default holds when controlling for other loan characteristics and housing price. Moreover, all of the results hold after using propensity score matching, based on borrowers' characteristics (including income, race, sex) and loan characteristics (e.g. original loan balance, location of property, origination year, etc.) for different loan types. These results are therefore the first direct support for the claim that the mortgage default decisions of borrowers is related to their different time preferences.

The rest of this chapter is organized as follows: in sections 2.2, the relationship between this chapter and the prior literature is presented; in section 2.3, following the original work of Phelps and Pollak (1968) (later employed by Laibson (1994, 1997, 1998) and other papers), a typical form of present-biased preference (i.e., quasi-hyperbolic) is presented; time preferences and mortgage choices are illustrated in section 2.4; section 2.5 outlines and describes the data used in this chapter; section 2.6 discusses the default behavior for different mortgage types using a logistic model; and section 2.7 concludes this chapter.

2.2 Relationship to the Prior Literature

This chapter contributes to several strands of the existing literature. First, it contributes to the broader literature on mortgage default. It presents an alternative theory to explain the origins of the unobserved heterogeneity of mortgage borrowers and how this unobserved heterogeneity affects borrowers' default decisions and behaviors. The "frictionless" market is an ideal assumption, and transaction costs have a significant impact on the likelihood of individuals' decisions to strategically default or prepay. Foster and van Order (1984) found that borrowers would not default "ruthlessly", and exercised the put option of default if the value of their house fell below the mortgage value by an amount equal to the net transaction costs. Therefore, we argue that a pure option-based theory cannot fully explain mortgage default behavior, and the impact of transaction costs on the mortgage default decision is important (Cunningham and Hendershott 1984; Foster and Van Order 1984, 1985; Vandell and Thibodeau 1985; Quigley and Van Order 1991; Lekkas, Quigley, and Van Order 1993).

In addition to the above studies that conjecture the existence of transaction costs, some other studies try to test the importance of transaction costs and incorporate them into default risk modelling. For example, Lekkas et al. (1993) and Quigley and Van Order (1995) explicitly tested the 'frictionless' models, and validated the importance of transaction costs besides equity position.[8] Deng et al. (2000), Deng, Pavlov, and Yang (2005), and Clapp, Deng, and An (2006) stressed the importance of borrower and spatial heterogeneity associated with transaction costs.[9] Kau, Keenan, and Kim (1993) and Kau and Slawson (2002) included both transaction costs and suboptimal termination into mortgage pricing models.

At the same time, transaction costs are complicated and differ across mortgage holders, creating significant unobserved heterogeneity among borrowers. The role of unobserved borrower heterogeneity in explaining mortgage termination has also attracted much attention. Richard and Roll (1989), Schwartz and

---

[8]Clauretie (1987) and Hendershott and Schultz (1993) explicitly included non-equity variables, like costs of foreclosure, unemployment rate and showed their significance for foreclosure rates.
[9]Other examples incorporating market imperfections to "frictionless" models include Vandell et al. (1993), Archer et al. (2002), Van Order and Zorn (2000), Clapp et al. (2001), Pavlov (2001), Calhoun and Deng (2002), Goldberg and Harding (2003), etc.

Torous (1989), and Archer, Ling, and McGill (1996) all suggested the existing of ad hoc variables in analysis of pools, and addressed heterogeneity within mortgage pools. Stanton (1995, 1996) extended previous research to handle heterogeneity between mortgage pools. Deng et al. (2000) considered the issue of unobserved heterogeneity in the context of hazard modelling, and they explicitly accounted for the unobserved heterogeneity among borrowers by adding discretely distributed mass point mixed hazard. This research has been followed by Clapp, Deng and An (2006). Although transaction costs and unobserved heterogeneity have been discussed extensively in empirical studies, there is no unifying theory to explain the underlying unobserved heterogeneity of borrowers.

This chapter also adds to the literature on individual different time preferences, which has been addressed both in psychology and behavioral economics. Past research has documented that individual differences in time preference are an important predictor in many life choices such as gym contracts (DellaVigna and Malmendier 2006), smoking (Gul and Pesendorfer 2007), body-mass index (Smith, Bogin, and Bishai 2005; Courtemanche and Carden 2011), savings for retirement (Carroll et al. 2009), and credit card debt (Laibson, Repetto, and Tobacman 2007; Chabris, Laibson, and Schuldt 2008; Reimers et al. 2009; Meier and Sprenger 2010; Kuchler, 2013). In addition, Krusell, Kuruscu, and Smith (2010) analysed the effects of present bias on optimal taxation. DellaVigna and Paserman (2005) used present bias preference to explain individual job search behaviour. However, researchers have not studied the effects of present bias on mortgage choice and default, and relatively few papers on mortgage choice and default have distinguished between naïve and sophisticated individuals. In this chapter, heterogeneous time preferences among borrowers, as indicated by their mortgage choices, are used to explain the default behavior of present-biased borrowers. Different mortgage types are also used to differentiate naïve and sophisticated borrowers.

2.3 Present-Biased Preference

Traditional inter-temporal preference models in economics have captured the impatience of agents by using exponential discounting. This approach explicitly assumes that preferences are inter-temporally consistent. However, this standard economic assumption may not be applicable in all instances when we are considering trade-offs between two future moments. Specifically, individuals with present-biased preference tend to give relatively more weight to nearer moments in the future as they get closer, and their inter-temporal preferences are time inconsistent (O' Donoghue and Rabin 1999a).

One way of modelling present-biased preference is to use "quasi-hyperbolic discounting" or "(β, δ)-preference". This method was originally developed by Phelps and Pollak (1968), and was later employed by Laibson (1994, 1997, 1998) to capture self-control problems within an individual. This method is widely used in the literature (e.g. O' Donoghue and Rabin 1999a, 1999b, 1999c, 2001; Carrillo 1999; Fischer 2001) and will be used in this chapter.

Let u_t be the instantaneous utility the borrower gets in period t, and $U(u_t, u_{t+1}, u_{t+2}, \dots, u_T)$ be the borrower's inter-temporal preference function from the perspective of period t. Borrowers are assumed to have quasi-hyperbolic preference. Time is divided into two periods: the present period (t) and all future periods (beginning from t+1 to T). The inter-temporal preference function for borrowers with present-biased preference can be expressed as:

(2.1) $$U(u_t, u_{t+1}, u_{t+2}, \dots, u_T) = \delta^t u_t + \beta \sum_{\tau=t+1}^{T} \delta^\tau u_\tau$$

$$\text{for } t \in [1, T];\ 0 < \delta \le 1;\ 0 < \beta \le 1$$

As in the standard exponential discounting model, the parameter δ represents the "time-consistent" long-run discounting factor. In this inter-temporal preference model, an additional parameter β is added into the standard time-consistent model for the future period to capture an individual's "bias for the present" – i.e., the agent's preference for the current over all future periods. There are two types of β: $\beta = 1$ and $0 < \beta < 1$.

For $\beta = 1$,

(2.2) $$U(u_t, u_{t+1}, u_{t+2}, \dots, u_T) = \sum_{\tau=t}^{T} \delta^\tau u_\tau$$

The inter-temporal preference function reduces to a standard exponential discounting utility function with time-consistent inter-temporal preference (the discrete version). Under this time preference, borrowers treat the present period and all future periods the same.

For $0 < \beta < 1$,

(2.3) $$U(u_t, u_{t+1}, u_{t+2}, \dots, u_T) = \delta^t u_t + \beta \textstyle\sum_{\tau=t+1}^{T} \delta^\tau u_\tau$$

This function parsimoniously captures present-biased preference, and greater weight is assigned to the present relative to the future. The β-parameter in this model thus fully captures the dynamic-inconsistency suggested by present-biased preference (O' Dognohue and Rabin 1999a, 1999b 1999c).

If time-inconsistent preference is assumed, an individual at each time period is modelled as a separate agent who maximizes utility according to her current preference, while her "future selves" will control her future behavior based on the prevailing preferences in the future (O'Donoghue and Rabin 1999a). An important question following this assumption is: what does an individual believes about her future selves' preferences? A crucial insight from the present-biased preference perspective is the distinction between naïve and sophisticated individuals. A sophisticated individual is fully aware of her future self-control problems, and knows her future selves' preferences exactly, even though they may differ from those of the current self. In contrast, a naïve individual does not anticipate her future procrastination, and is thus fully unaware of her future self-control problems. This inclines her to believe that her future preferences will be identical to her current ones. Under such a distinction, choices of naïve individual and sophisticated individual are different.

2.4 Time Preferences and Mortgage Choice

Selecting a mortgage is a consequential consumer choice that highlights the role of time preferences in determining outcomes.[10] While mortgage are often complex and differ along many dimensions, they can be broadly classified into

[10] Carroll et al. (2009) modelled the optimal policies of 401(k) saving for present biased consumers; DellaVigna and Malmendier (2004) studied the present biased consumers' contract choice among health club, and concluded that consumer's preferences among contract were important; Prelec and Loewenstein (1998) illustrated how payment and consumption events can be optimally timed and linked.

two main categories based on their repayment structure: back-loaded mortgages and front-loaded mortgages. A second dimension of interest is the length of repayments. Mortgage contracts typically involve repayment periods of 30-year, but can also be structured for shorter periods, such as 10-year. The selection of a particular payment structure is an indication of a borrower's intertemporal preferences

A fixed-rate mortgage (FRM) is a fully amortizing mortgage loan where the interest rate on the note remains the same through the term of the loan, as opposed to "floating" loans where the interest rate may adjust. As a result, the payment amounts and loan duration of an FRM are fixed and the person who is responsible for paying back the loan benefits from a consistent, single payment and the ability to plan a budget based on this fixed cost. The constant discounting for fixed-rate mortgage implies that a person's intertemporal preferences are time-consistent, which means that any decision that the individual makes for himself in advance will remain valid as time advances, and later preferences "confirm" earlier preferences. Therefore, borrowers with time-consistent preference (standard exponential discounting) will choose fixed-rate mortgages.

An interest-only loan is a loan in which, for a set term, the borrower pays only the interest on the principal balance, with the principal balance unchanged. At the end of the interest-only term, the borrower may enter an interest-only mortgage, pay the principal, or (with some lenders) convert the loan to a principal and interest payment (or amortized) loan at his/her option. Mortgages such as interest-only loans are particularly appealing to present-biased individuals because they present lower upfront costs in return for greater later costs. In addition, with the absence of self-control, interest-only loans are more likely to be selected by naïve borrowers who are fully unaware of their future self-control problems, minimize their up-front costs and postpone the payment on their mortgages.

In contrast, 'sophisticated' borrowers who have present-biased preference, or suffer from short-term temptations and are aware of the consequences, are likely to prefer to control themselves from temptation and behave more rationally. When selecting a mortgage, sophisticated borrowers may be worried about the minimal up-front costs and the corresponding future over-payments, and may refrain from them so as to induce themselves to resist temptation in the future. This means that, unlike naïve borrowers, sophisticated borrowers will not choose interest-only loans. However, they will also not choose fixed-rate loans, which would be preferred by time-consistent borrowers. An adjustable-rate mortgage (ARM) is a mortgage loan where the interest rate on the note is periodically adjusted based on an index which reflects the cost to the lender of borrowing on the credit markets. Adjustable-rate mortgage loans are similar to interest-only loans, in that they allow borrowers to enjoy minimal up-front costs, and differ from fixed-rate mortgage loans, in that loan repayments are not a consistent amount. Therefore, sophisticated borrowers with self-control will be more likely to choose an adjustable-rate mortgage.

2.5 Data

Two main sources of data are used in this chapter: individual loan-level mortgage data, from BlackBox Logic (BBL), and the database of home loan applications and originations collected by the Home Mortgage Disclosure Act (HMDA).

BlackBox Logic (BBL) is a private company that provides a comprehensive, dynamic dataset with information about twenty-one million privately securitized subprime, Alt-A, and prime loans in the United States. These loans account for about ninety percent of all privately securitized mortgages. The BlackBox data, which is obtained from mortgage services and securitization trustees, includes static information taken at the time of the origination of mortgages, such as the mortgage contract date, original loan amount, the initial loan-to-value ratio, borrowers' FICO credit scores, mortgage service name, mortgage contract interest rate, mortgage term, interest rate type, state, region, and major metropolitan area in which the property is located. In addition, the BlackBox data also include dynamic data on monthly payments, mortgage balances, current loan to value ratio, and delinquency status.

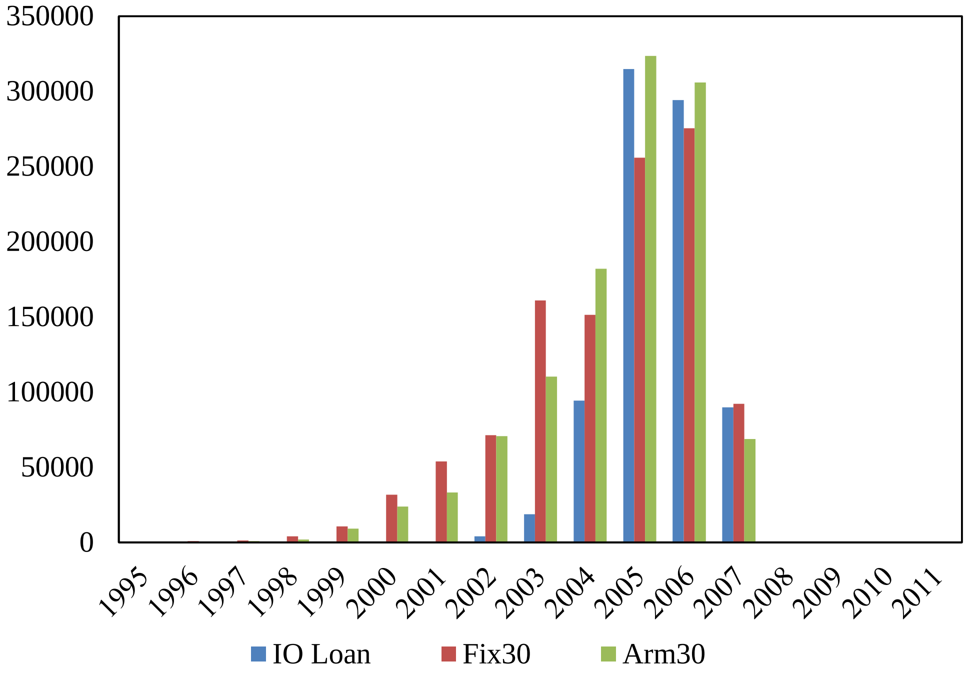
The HMDA database is available at the loan application level.[11] It is an annual database that contains each applicant's final status (denied/approved/originated), purpose of borrowing (home purchase/refinancing/home improvement), loan amount, borrowers' attributes (race, gender, income, and home ownership status), and also (in the case of originated loans) whether the loan was sold to the secondary market within the year. In addition, the location of property is clearly recorded in the HMDA database.

[11] The Home Mortgage Disclosure Act (HMDA), enacted by Congress in 1975 and implemented by the Federal Reserve Board, requires lending institutions to report public loan data. The lending institutions mainly include banks, savings associations, credit unions, and other mortgage lending institutions.

The analysis in this chapter is confined to first-lien mortgage loans issued between 1995 and 2011, and includes those loans that were either closed or still active at the third quarter of 2012. The analysis is confined to 5-year interest-only loans, 10-year interest-only loans, 30-year fixed-rate loans and 30-year adjustable-rate loans. After removing mortgages with incomplete information on LTV ratio, original loan balance, FICO score and other key information, the final sample includes 3,058,413 individual mortgages.

Figure 2-1 shows the distribution of loan origination over the years. Generally, the number of all kinds of loans grew tremendously between 2001 and 2006. In 2005, the number of originations of interest-only loans and adjustable-rate loans reached a peak. The number of fixed-rate loans originations peaked in 2006. Before 2003, the origination of fixed-rate loans dominated the loan market, and this changed from 2004. The origination of interest-only and 30-year adjustable-rate loans expanded very quickly from 2004. Moreover, the number of interest-only loans grew at a faster rate than both fixed-rate loans and adjustable-rate loans. After the financial crisis, the origination of all kinds of loans decreased sharply.

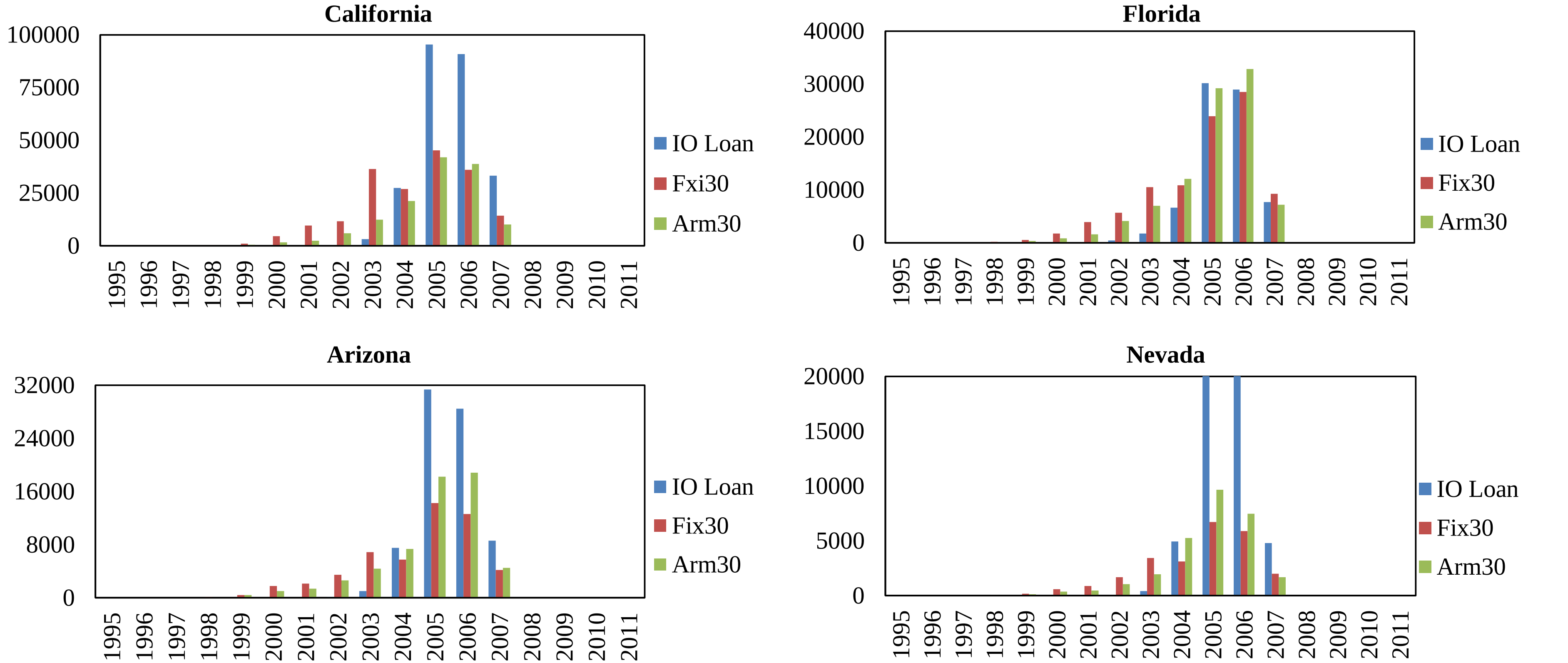
Figure 2-1 Frequency Distribution by Origination Year: Full Sample



Note: This figure shows the frequency distribution of all kinds of loans originations for the full sample. Y-axis measures the frequency of origination, while X-axis measures the year. Three loan types are included in this sample: interest-only loans, 30-year fixed-rate loans and 30-year adjustable-rate loans.

Figure 2-2 focuses on the loan origination growth in four states (i.e., California, Florida, Arizona and Nevada). Consistent with the dramatic growth as shown in Figure 2-1, the origination of all kinds of loans grew tremendously between 2001 and 2006 in these four states. In addition, the origination of interest-only loans increased particularly fast in California (CA), Arizona (AZ) and Nevada (NV). In these three states, the origination of interest-only loans in 2005 and 2006 was more than twice the number of 30-year fixed-rate loans. Consistent with Figure 2-1, the number of originations for all kinds of loans dropped sharply since the financial crisis began.

Figure 2-2 Frequency Distribution by Origination Year: Four States



Note: This figure shows the frequency distribution of all kinds of loan originations for four states: California (CA), Florida (FL), Arizona (AZ) and Nevada (NV). Three loan types are included in this sample: interest-only loans, 30-year fixed-rate loans and 30-year adjustable-rate loans. Y-axis measures the frequency of origination, while X-axis measures the year.

Table 2-2 show the summary statistics of the BlackBox (BBX) dataset. Information on three types of first-lien mortgage loan originations between 1995 and 2011 are kept: interest-only loans (5-year and 10-year), 30-year fixed-rate loans and 30-year adjustable-rate loans. 26.7% of the loans are interest-only loans, with 10.62% being 5-year interest-only loans and 16.08% 10-year interest-only loans. 36.31% of the loans are 30-year fixed-rate loans, and 36.99% are 30-year adjustable-rate loans. Borrowers have an average FICO score of 663.52, and borrowed up to 79.15% of the property value (LTV) in the sample.

Columns (2) to (4) show the summary statistics for each loan type separately. Nearly 40% of interest-only loans are 5-year interest-only loans. Loans and borrowers have distinct characteristics for each loan type. The average amount borrowed is the highest among interest-only loans, nearly two times the amount borrowed for 30-year adjustable-rate loans. The FICO scores for interest-only loans are the highest with an average of 699.496. Borrowers who opted for 30-year adjustable-rate loans have the lowest FICO score of 617.575. The average amount borrowed for interest-only loans and 30-year fixed-rate loans is similar, and is up to 77% of the property's value. However, the average amount borrowed for 30-year adjustable-rate loans is much higher than other types of loans, up to 82.5% of the property's value.

Table 2–2 Summary Statistics of Variables: Full Sample

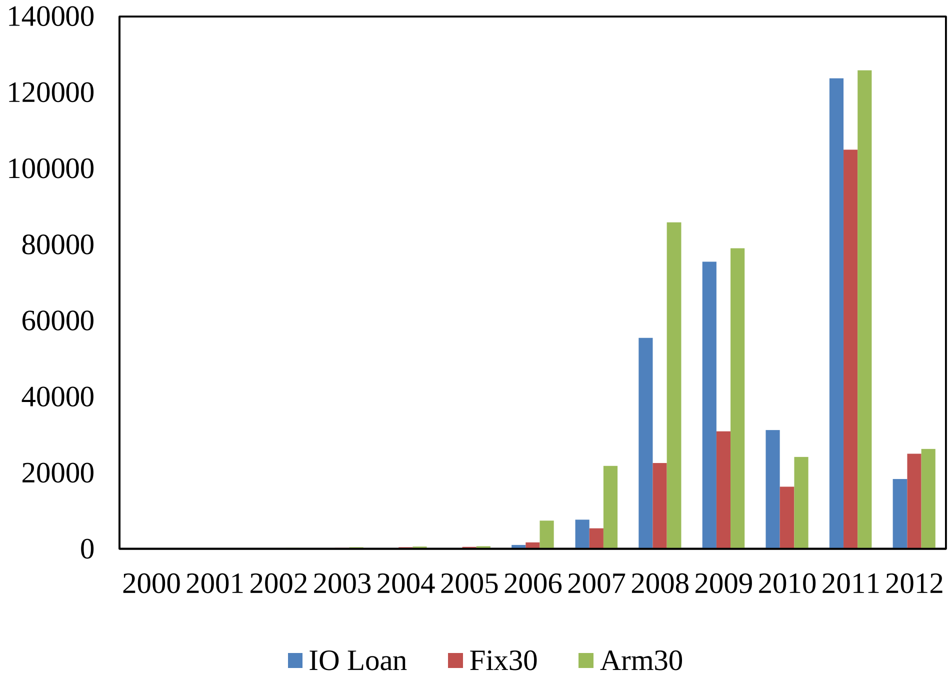
	Original total	IO Loans	FIX30	ARM30
IO5	10.62%	39.77%		
IO10	16.08%	60.23%		
FIX30	36.31%		100%	
ARM30	36.99%			100%
Current Interest rate	7.001	5.969	6.984	7.790
Original Loan Balance	225143.32	312080.52	216238.08	171130.41
FICO Score	663.52	699.496	683.866	617.575
OrigLTVRatioCalc	79.150	77.176	77.187	82.502
ownerocc	84.44%	81.68%	81.36%	89.45%
low_no_doc	45.22%	60.37%	45.88%	33.62%
subprime	29.34%	8.85%	18.89%	54.38%
Duration	56.606	54.444	62.426	52.452
LOG_MHPI	5.168	5.203	5.150	5.161
Sample size	3,058,413	816,569	1,110,638	1,131,206

Note: This table presents the aggregate-level summary statistics of BlackBox dataset. The sample only include 5-year interest-only loans, 10-year interest-only loans, 30-year fix-rate loans and 30-year adjustable-rate loans. Comprarsion of the average values of variables by full sample, interest-only loans, 30-year fix-rate loans and 30-year adjustable-rate loans are presented. '*IO5'* is represetation of the 5-year interest-only loans, takes the value of one for 5-year interest-only loans and zero for others; '*IO10'* is represetation of the 10-year interest-only loans, takes the value of one for 10-year interest-only loans and zero for others; '*FIX30'* is represetation of the 30-year fix-rate loans, takes the value of one for 30-year fix-rate loans and zero for others; '*ARM30'* is represetation of the 30-year adjustable-rate loans, takes the value of one for 30-year adjustable-rate loans and zero for others; '*Original Loan Balance*' is defined as the amount of principal on the closing date of the mortgage; '*FICO Score'* refers to the Fair Issacs borrowers score at the time of loan origination. '*OrigLTVRatioCalc'* refers to the ratio of the original loan amount to the property value at loan origination; '*Ownerocc'* takes the value of one if the property type is owner occupied and zero for others; '*low_no_doc'* takes the value of one if the documentation type of the loan is low or no documentation and zero for others; *'Subprime'* equals to one if loan is subprime and zero for others; '*Duration'* is the duration of the loans in month, which is defined as the elapsed time from origination to the end of the sample period or to the first classification as being prepaid or delinquent at least 60 days; '*LOG_MHPI'* is the logarithm for quarterly FHFA house price index.

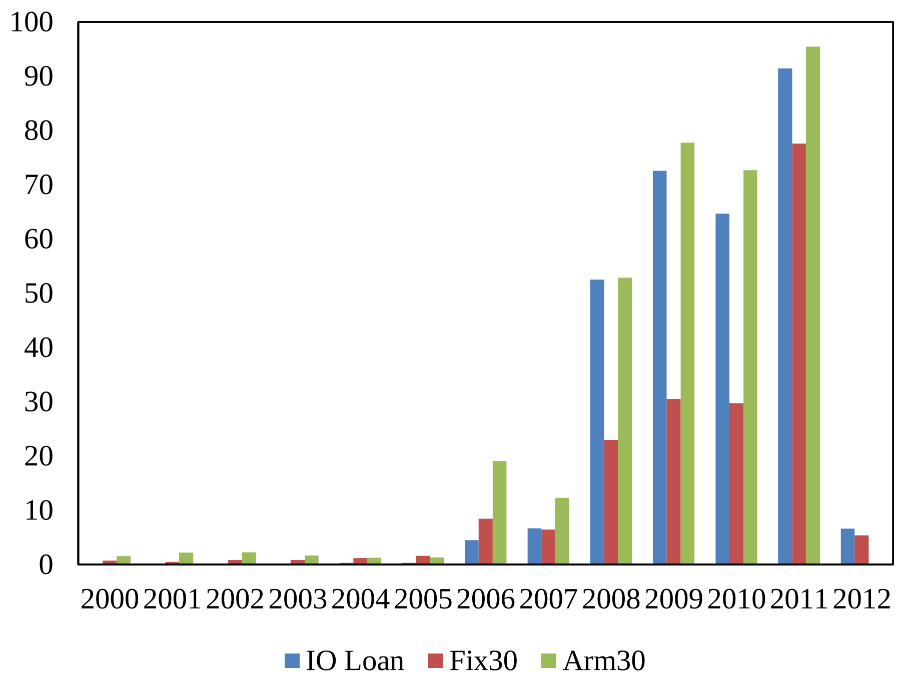
Figure 2-3 suggests the default pattern for each type of loan over the period being studied. The default rate for each type of loan increased dramatically after the recent financial crisis, especially for interest-only loans and 30-year adjustable-rate loans. Both the default frequency and default percentage reached their peaks in 2011 for all kinds of loans. In addition, it can be seen, from both the frequency and percentage of default distributions, that the default rate for interest-only loans and 30-year adjustable-rate loans is much higher than 30-year fix-rate loans from 2008 to 2011. Figure 2-4 suggests that the default patterns in the four states are consistent with the full sample. Particularly, the default frequency for interest-only loans after the recent financial crisis is much higher than the default frequency for others type of loans in CA, AZ and NV.

Figure 2-3 Default Distribution over Years: Full Sample

A. Frequency Distribution of Default by Year



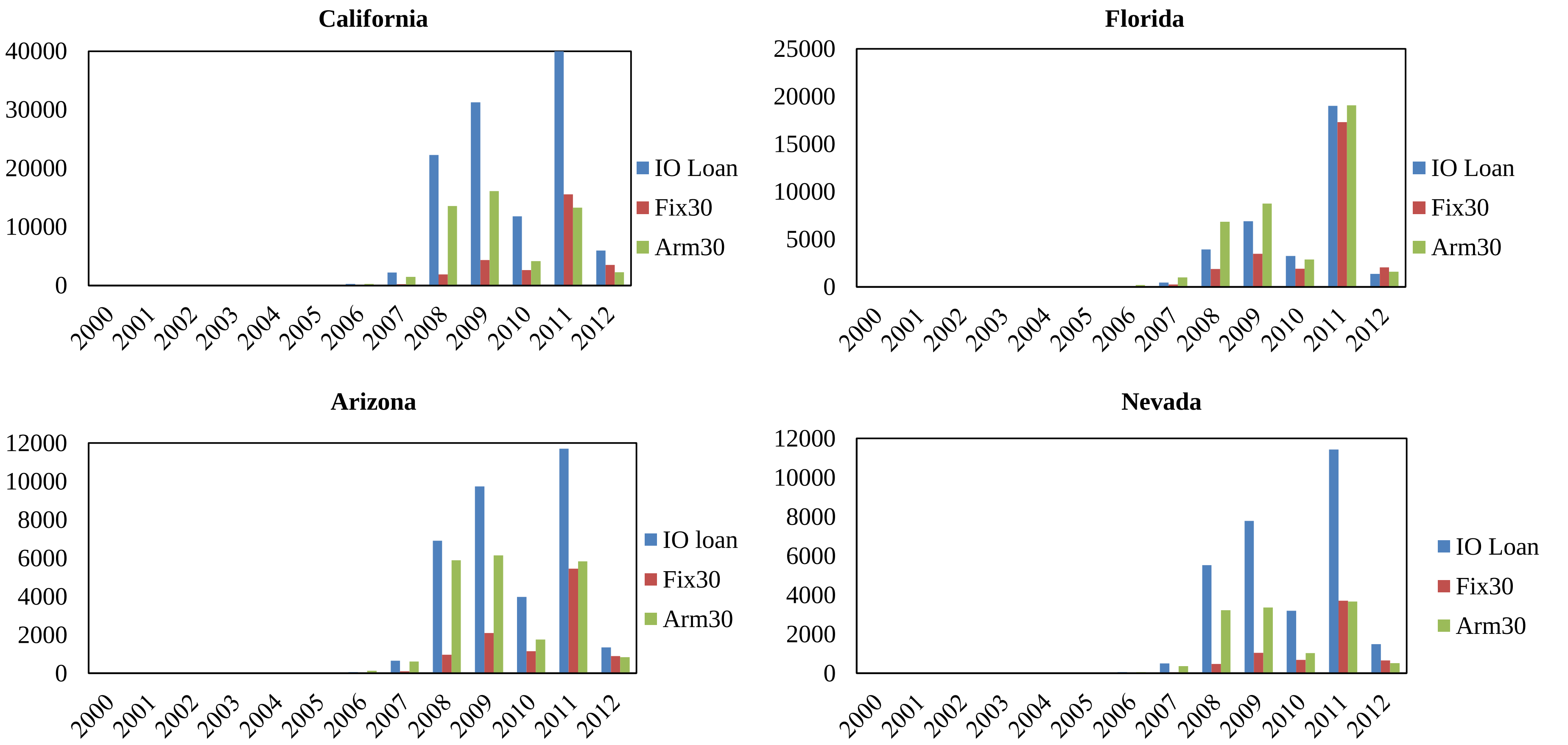
B. Percentage of Default by Year



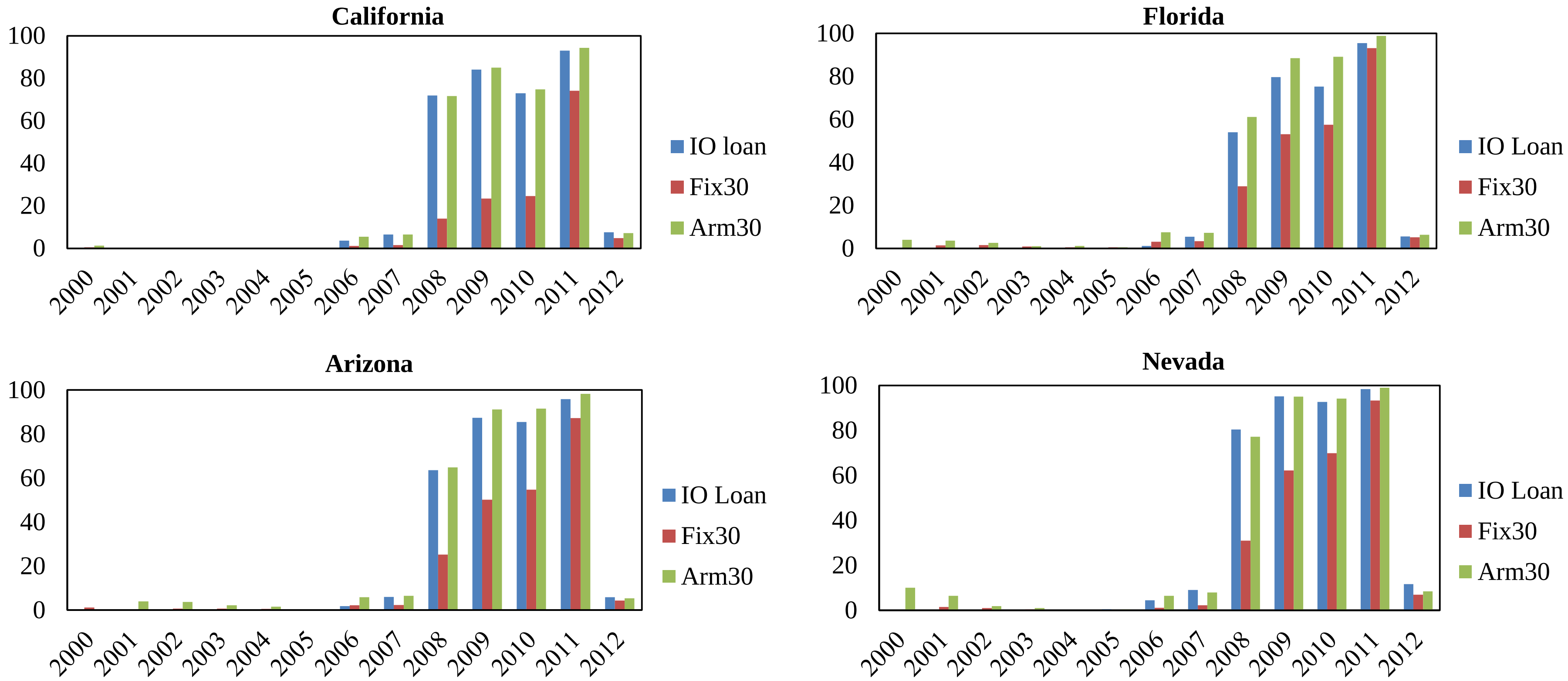
Note: This figure shows the frequency and percentage of default distribution for all kinds of loans over year. In panel A, Y-axis measures the frequency of default loans, while X-axis measures the year. In panel B, Y-axis measures the default rate, while X-axis measures the year.

Figure 2-4 Default Distribution over Years: Four States

A. Frequency Distribution of Default by Year



B. Percentage of Default by Year



Note: This figure shows the default frequency and percentage distribution of all kinds of loans by year for four states: California (CA), Florida (FL), Arizona (AZ) and Nevada (NV). In panel A, Y-axis measures the frequency of default loans, while X-axis measures the year. In panel B, Y-axis measures the default rate, while X-axis measures the year.

2.6 Empirical Results

2.6.1 Time Preferences and Mortgage Default

Logistic regression is employed to study the default behavior of interest-only loans and 30-year adjustable-rate loans, relative to 30-year fix-rate loans. Table 2-3 reports the regression coefficients and odds ratios in the full sample analysis. Consistent with existing findings on the determinants of the default behavior, owner-occupancy, lower LTV ratio, high FICO score and lower loan balance predict a lower default rate in general. In addition, lower or no documentation loans are risky, and their default probabilities are higher. Column (1) shows the regression results of interest-only loans relative to 30-year fixed-rate loans. It can be seen that 5-year interest-only loans are 41% more likely to default than 30-year fixed-rate loans after controlling other loan characteristics. The default probability for 10-year interest-only loans is 47.4% higher than that for 30-year fixed-rate loans, after controlling other loan characteristics.

The results in column (2) show the regression results for 30-year adjustable-rate loans relative to 30-year fixed-rate loans. 30-year adjustable-rate loans are 26.8% more likely to default compared with 30-year fixed-rate loans, after controlling for other loan characteristics. The last column shows the results in the full sample analysis, where the default probabilities for both interest-only loans and 30-year adjustable-rate loans relative to 30-year fixed-rate loans are presented. Compared with 30-year fixed-rate loans, the default probability of 5-year interest-only loan is 34.9% higher, and that for 10-year interest-only loans is 39.3% higher. 30-year adjustable-rate loans are 24.4% more likely to default relative to 30-year fixed-rate loans. Moreover, the default probability of 30-year adjustable-rate loans is lower than interest-only loans.

Table 2–3 Time Preferences and Mortgage Default: Full Sample

Variables	IO VS FIX30	ARM30 VS FIX30	IO & ARM30 VS FIX30
IO5	0.343***		0.300***
	(0.003)		(0.002)
	[1.408]		[1.349]
IO10	0.388***		0.331***
	(0.003)		(0.002)
	[1.474]		[1.393]
ARM30		0.238***	0.219***
		(0.003)	(0.003)
		[1.268]	[1.244]
ownerocc	-0.188***	-0.211***	-0.187***
	(0.003)	(0.002)	(0.002)
	[0.828]	[0.810]	[0.830]
low_no_doc	0.317***	0.262***	0.306***
	(0.003)	(0.003)	(0.002)
	[1.373]	[1.299]	[1.358]
subprime	-0.112***	-0.167***	-0.172***
	(0.004)	(0.004)	(0.003)
	[0.894]	[0.846]	[0.842]
OrigLTVRatioCalc	0.751***	0.638***	0.689***
	(0.003)	(0.003)	(0.003)
	[2.118]	[1.892]	[1.992]
log_FICO Score	-1.015***	-0.989***	-1.041***
	(0.004)	(0.004)	(0.004)
	[0.363]	[0.372]	[0.353]
log _ Original Loan Balance	-0.233***	-0.195***	-0.203***
	(0.004)	(0.003)	(0.003)
	[0.792]	[0.823]	[0.816]
LOG_MHPI	-0.436***	-0.351***	-0.398***
	(0.004)	(0.003)	(0.003)
	[0.647]	[0.704]	[0.671]
log_duration	0.613***	0.984***	0.885***
	(0.010)	(0.009)	(0.008)
	[1.846]	[2.674]	[2.422]
Observations	1,927,207	2,241,844	3,058,413
Pseudo R-Squared	0.6892	0.6504	0.6788

Note: This table shows results of logistic regression analysis for BlackBox dataset. The sample only includes 5-year interest-only loans, 10-year interest-only loans, 30-year fix-rate loans and 30-year adjustable-rate loans. The dependant varaible is ‘default; takes the value of one for default loans and zero for others. The definitions of the independent variables are shown in Table 2-2. State, current year and origination year fixed effects are included in the regression but not reported. Standard errors are reported in parentheses, and estimated odds ratios for Logit regression are reported in brackets.
*Significant at the 10% level;
** Significant at the 5% level;
*** Significant at the 1% level.

Table 2-4 presents the sub-sample default analysis in California, Florida, Arizona and Nevada respectively. As shown in Figure 2-2 and Figure 2-4, the patterns of origination and default for each loan type in these four states are similar to the full sample. Consistent with the results from the full sample, the average default rates for interest-only loans and 30-year adjustable-rate loans are higher relative to 30-year fixed-rate loans, and the default probability of 30-year adjustable-rate loans is lower than interest-only loans. However, compared with the results in Table 2-3, the default probability of interest-only loans relative to 30-year fixed-rate loans in these four states is higher than the general results, and the default probability of 30-year adjustable-rate loans relative to 30-year fixed-rate loans in these four states is lower than the general results. In particular, in California, borrowers of 5-year interest-only loans are around 58.6 percentage points more likely to default than those who selected 30-year fixed-rate loans, and 10-year interest-only loans borrowers are around 64.5 percentage points more likely to default than those who selected 30-year fixed-rate loans. On the other hand, in California, 30-year adjustable-rate loans borrowers are around 25.6 percentage points more likely to default than those who have selected 30-year fixed-rate loans.

Table 2–4 Time Preferences and Mortgage Default: Four States

Panel A Logistic Regression: State of California (CA)			
	IO VS FIX30	ARM30 VS FIX30	ARM30 & IO VS FIX30
IO5	0.461***		0.423***
	(0.008)		(0.007)
	[1.586]		[1.526]
IO10	0.498***		0.442***
	(0.008)		(0.007)
	[1.645]		[1.556]
ARM30		0.228***	0.160***
		(0.009)	(0.008)
		[1.256]	[1.173]
Observations	437,487	322,103	573,012
Pseudo R-Squared	0.8000	0.7900	0.8063

Panel B Logistic Regression: State of Florida (FL)			
	IO VS FIX30	ARM30 VS FIX30	ARM30 & IO VS FIX30
IO5	0.258***		0.206***
	(0.010)		(0.008)
	[1.294]		[1.229]
IO10	0.328***		0.254***
	(0.011)		(0.009)
	[1.388]		[1.289]
ARM30		0.178***	0.141***
		(0.011)	(0.010)
		[1.195]	[1.151]
Observations	171,346	191,050	266,849
Pseudo R-Squared	0.7961	0.7992	0.8019

Panel C Logistic Regression: State of Arizona (AZ)			
	IO VS FIX30	ARM30 VS FIX30	ARM30 & IO VS FIX30
IO5	0.365***		0.336***
	(0.0138)		(0.012)
	[1.441]		[1.399]
IO10	0.510***		0.436***
	(0.0143)		(0.013)
	[1.665]		[1.154]
ARM30		0.248***	0.190***
		(0.015)	(0.014)
		[1.281]	[1.209]
Observations	128,980	110,668	187,854
Pseudo R-Squared	0.8084	0.7917	0.8060

Panel D Logistic Regression: State of Nevada (NV)			
	IO VS FIX30	ARM30 VS FIX30	ARM30 & IO VS FIX30
IO5	0.285***		0.269***
	(0.019)		(0.017)
	[1.329]		[1.308]
IO10	0.452***		0.421***
	(0.020)		(0.018)
	[1.572]		[1.523]
ARM30		0.159***	0.108***
		(0.022)	(0.019)
		[1.173]	[1.114]
Observations	75,076	52,577	103,097
Pseudo R-Squared	0.8236	0.8167	0.8241

Note: This table shows results of logistic regression analysis for four representative states, i.e., California, Florida, Arizona and Nevada. The dependant varaible is ‘default; takes the value of one for default loans and zero for others. We do not report the entire list of control variables, and refer to Table 2-3 for the full lists. The definitions of the independent variables are shown in Table 2-2. State, current year and origination year fixed effects are included in the regression but not reported. Standard errors are reported in parentheses, and estimated odds ratios for Logit regression are reported in brackets.
*Significant at the 10% level;
** Significant at the 5% level;
*** Significant at the 1% level.

Table 2-5 presents the sub-sample default analysis by setting the financial crisis as a break point. Specifically, the sample is divided into two parts: loans that terminated before the financial crisis and loans that terminated after financial crisis.[12] Generally speaking, for both before and after financial crisis sample, all results are consistent with the results in the full sample. Before the financial crisis, 5-year interest-only loans were 14.5% more likely to default than 30-year fixed-rate loans, after controlling for other loan characteristics. The default probability for 10-year interest-only loans is only 8.9% higher than that for 30-year fixed-rate loans, after controlling other loan characteristics.

[12] For simplicity, the financial crisis is defined as beginning at the start of 2009. Therefore, the two parts of the sample are: loans that terminate before the end of 2008 and loans that terminate after the beginning of 2009.

The results changed dramatically after the financial crisis. 5-year interest-only loans were 44% more likely to default than 30-year fixed-rate loans, after controlling for other loan characteristics. The default probability of 10-year interest-only loans was 50.6% higher than 30-year fixed-rate loans. However, the results for 30-year adjustable-rate loans compared to 30-year fixed-rate loans were the same. This divergence in the findings can be explained by the higher probability for back-loaded mortgages that minimize up-front costs to go “underwater” and default following negative home price shocks.

Table 2–5 Time Preferences and Mortgage Default: Financial Crisis Break Point

	IO VS FIX_30	ARM_30 VS FIX_30	ARM_30 & IO VS FIX_30
Terminate before Financial Crisis			
IO5	0.135***		0.113***
	(0.009)		(0.006)
	[1.145]		[1.120]
IO10	0.086***		0.040***
	(0.010)		(0.007)
	[1.089]		[1.041]
ARM30		0.114***	0.111***
		(0.0070)	(0.007)
		[1.120]	[1.117]
Observations	398,861	623,388	768,469
Pseudo R-Squared	0.2292	0.2702	0.2457
Terminate after Financial Crisis			
IO5	0.365***		0.320***
	(0.003)		(0.002)
	[1.440]		[1.377]
IO10	0.410***		0.357***
	(0.003)		(0.003)
	[1.506]		[1.429]
ARM30		0.239***	0.205***
		(0.003)	(0.003)
		[1.270]	[1.227]
Observations	1,528,346	1,618,456	2,289,944
Pseudo R-Squared	0.6962	0.6665	0.6909

Note: This table shows results of logistic regression analysis for financial crisis break point. For simplify, here the time of financial crisis defined as the beginning of 2009. Therefore, the two parts of sample are: loans terminate before the end of 2008 and loans terminate after the beginning of 2009.The dependant variable is ‘default; takes the value of one for default loans and zero for others. We do not report the entire list of control variables, and refer to Table 2-3 for the full lists. The definitions of the independent variables are shown in Table 2-2. State, current year and origination year fixed effects are included in the regression but not reported. Standard errors are reported in parentheses, and estimated odds ratios for Logit regression are reported in brackets.

*Significant at the 10% level;
** Significant at the 5% level;
*** Significant at the 1% level.

2.6.2 Robustness Analysis: Propensity Score Matching

The original loan size and other observable attributes of each loan type are systematically different (see Table 2-2), and directly comparing their default may be misleading with an unbalanced sample. Therefore, propensity score matching (PSM) is used to obtain a more homogeneous sample for each comparison sample to mitigate the potential bias.[13] Specially, a one-to-one match for each treatment group based on the original loan balance, origination year, location of property (MSA level) and other loan and borrower characteristics is performed.

Table 2-6 reports the summary statistics of the matched sample. Although propensity score matching is not able to entirely eliminate the difference in loan and borrower characteristics of the interest-only loans and 30-year adjustable-rate loans relative to the 30-year fixed-rate loans, the gap between those observables is greatly reduced across these three loan types after matching. Firstly, interest-only loans (the treatment group) are matched with 30-year fixed-rate loans (the control group) to mitigate the potential bias. In the matched sample, 22.44% the loans are 5-year interest-only loans and 27.56% are 10-year interest-only loans. The statistics of all the variables are very similar after matching for interest-only loans and 30-year fixed-rate loans. Second, the summary statistics are shown by matching 30-year adjustable-rate loans (the treatment group) with 30-year fixed-rate loans (the control group). Lastly, both interest-only loans and 30-year adjustable-rate loans (the

[13] The BlackBox dataset has fewer borrowers' characteristics than the HMDA dataset. Therefore, firstly, the HMDA dataset is matched with the BlackBox dataset to obtain more information about mortgage borrowers, including borrowers' race, sex, income, and home ownership status. Then the propensity score matching is carried out based on the combined information from both datasets.

treatment group) with 30-year fixed-rate loans (the control group) are matched. After matching, 22.08% of the loans are interest-only loans, with 8.54% having a duration of 5-year and 13.54% being 10-year long, and 27.64% of the loans are 30-year adjustable-rate loans. In the treatment group, 44.72% of the loans are interest-only loans, with 17.08% being 5-year interest-only loans, 27.65% being 10-year interest-only loans, and 55.28% of the loans being 30-year adjustable-rate loans.

Table 2–6 Summary Statistics of Variables: Propensity Score Matched Sample

Matched Group	IO loans matched with FIX30			ARM30 matched with FIX30			IO loans and ARM30 matched with FIX30		
Variables	Matched total	IO Loans	Fix30	Matched total	Fix30	ARM30	Matched total	IO loans and Arm30	Fix30
IO5	22.44%	44..88%					8.54%	17.08%	
IO10	27.56%	55.12%					13.54%	27.65%	
ARM30				50%		100%	27.64%	55.28%	
Current Interest rate	6.311	5.943	6.777	7.388	7.304	7.473	6.889	6.779	6.997
Original Loan Balance	253257.42	260233.81	246281.03	174980.54	173859.90	176101.17	205232.94	204689.88	205776.01
FICO Score	694.293	694.035	694.552	644.457	644.736	644.178	673.229	672.690	673.768
OrigLTVRatioCalc	77.304	77.356	77.253	81.168	81.239	80.980	78.672	78.561	78.768
ownerocc	85.81%	85.85%	85.76%	88.33%	88.79%	87.86%	85.65%	85.38%	85.92%
low_no_doc	53.84%	53.88%	53.80%	37.70%	37.13%	38.26%	43.56%	43.21%	43.90%
subprime	11.18%	11.35%	11.01%	34.20%	34.23%	34.17%	23.42%	23.98%	22.85%
Duration	58.259	55.896	60.621	58.063	62.828	53.299	31.361	55.893	62.829
LOG_MHPI	5.183	5.202	5.165	5.154	5.138	5.171	5.164	5.176	5.151
Sample size	625,444	312,722	312,722	671,158	335,579	335,579	1,008,224	504,112	504,112

Note: This table presents the aggregate-level summary statistics of BlackBox dataset after Propenstiy Score matching. The sample only include 5-year interest-only loans, 10-year interest-only loans, 30-year fix-rate loans and 30-year adjustable-rate loans. Comprarsion of the average values of the variables by full sample, interest-only loans, 30-year fix-rate loans and 30-year adjustable-rate loans are presented. Definition of variables are the same as in Table 2-2.

Table 2-7 presents the results of the logistic regression analysis on the default behavior of interest-only loans and 30-year adjustable-rate loans relative to the 30-year fixed-rate loans in the matched sample. The findings are broadly consistent with those in Table 2-3: in the matched sample, the average default rate of interest-only loans and 30-year adjustable-rate loans is higher relative to that for 30-year fixed-rate loans, and the default probability for 30-year adjustable-rate loans is lower than that for interest-only loans.

Column (1) shows the regression results of interest-only loans relative to 30-year fixed-rate loans after matching. It can be seen that 5-year interest-only loans are 49.4% more likely to default than 30-year fixed-rate loans after controlling for other loan characteristics. In addition, the default probability of 10-year interest-only loans is 44.8% higher than 30-year fixed-rate loans after controlling for other loan characteristics. The results in column (2) have shown that 30-year adjustable-rate loans are 22.6% more likely to default compared to 30-year fixed-rate loans, after controlling for other loan characteristics and propensity matching. The last column shows the results in the full sample analysis, where the default probabilities of both interest-only loans and 30-year adjustable-rate loans relative to those for 30-year fixed-rate loans are presented. Compared with 30-year fixed-rate loans, the default probability of 5-year interest-only loans is 30.9% higher, and 31.6% higher for 10-year interest-only loans. The 30-year adjustable-rate loans are 21.7% more likely to default relative to 30-year fixed-rate loans. Moreover, the default probability of 30-year adjustable-rate loans is lower than the default probability of interest-only loans.

Table 2–7 Time Preferences and Mortgage Default:
Propensity Score Matched Sample

Variables	IO VS FIX30 (matched)	ARM30 VS FIX30 (matched)	IO & ARM30 VS FIX30 (matched)
IO5	0.401***		0.269***
	(0.005)		(0.004)
	[1.494]		[1.309]
IO10	0.370***		0.275***
	(0.005)		(0.004)
	[1.448]		[1.316]
ARM30		0.203***	0.197***
		(0.004)	(0.004)
		[1.226]	[1.217]
owneroc c	-0.229***	-0.218***	-0.237***
	(0.005)	(0.004)	(0.004)
	[0.795]	[0.804]	[0.789]
low_no_doc	0.312***	0.244***	0.281***
	(0.005)	(0.004)	(0.004)
	[1.366]	[1.276]	[1.324]
subprime	-0.138***	-0.112***	-0.147***
	(0.006)	(0.007)	(0.006)
	[0.871]	[0.894]	[0.863]
OrigLTVRatioCalc	0.745***	0.563***	0.695***
	(0.006)	(0.005)	(0.004)
	[2.107]	[1.756]	[2.003]
log_FICO Score	-0.982***	-0.829***	-1.027***
	(0.007)	(0.007)	(0.006)
	[0.375]	[0.437]	[0.358]
log _ Original Loan	-0.172***	-0.158***	-0.226***
Balance	(0.006)	(0.006)	(0.005)
	[0.842]	[0.854]	[0.798]
LOG_MHPI	-0.457***	-0.367***	-0.422***
	(0.007)	(0.006)	(0.005)
	[0.633]	[0.693]	[0.656]
log_duration	0.403***	0.856***	0.827***
	(0.018)	(0.016)	(0.014)
	[1.496]	[2.353]	[2.287]
Observations	625,444	671,158	1,008,224
Pseudo R-Squared	0.6868	0.6378	0.6456

Note: This table shows results of logistic regression analysis for BlackBox dataset after Propensity Score matching. The sample only includes 5-year interest-only loans, 10-year interest-only loans, 30-year fix-rate loans and 30-year adjustable-rate loans. The dependant varaible is ‘default; takes the value of one for default loans and zero for others. The definitions of the independent variables are shown in Table 2-2. State, current year and origination year fixed effects are included in the regression but not reported. Standard errors are reported in parentheses, and estimated odds ratios for Logit regression are reported in brackets.
*Significant at the 10% level;
** Significant at the 5% level;
*** Significant at the 1% level.

Table 2-8 presents the sub-sample default analysis in California, Florida, Arizona and Nevada respectively in the matched sample. The results are consistent with those in Table 2-4: the average default rates of interest-only loans and 30-year adjustable-rate loans are higher relative to 30-year fixed-rate loans, and the default probability of 30-year adjustable-rate loans is lower than that for interest-only loans. The default probability of interest-only loans is particular high in California, where borrowers of 5-year interest-only loans are around 70.1 percentage points more likely to default than those who selected 30-year fixed-rate loans, and borrowers of 10-year interest-only loans are around 56.2 percentage points more likely to default than those who selected 30-year fixed-rate loans.

Table 2–8 Time Preferences and Mortgage Default: Four States of Propensity Score Matched Sample

Panel A Logistic Regression: State of California (CA)			
	IO VS FIX30 (matched)	ARM30 VS FIX30 (matched)	ARM30 & IO VS FIX30 (matched)
IO5	0.531***		0.444***
	(0.013)		(0.011)
	[1.701]		[1.559]
IO10	0.446***		0.368***
	(0.013)		(0.011)
	[1.562]		[1.445]
ARM30		0.226***	0.155***
		(0.015)	(0.013)
		[1.254]	[1.168]
Observations	120,666	89,928	156,870
Pseudo R-Squared	0.7662	0.7798	0.7548

Panel B Logistic Regression: State of Florida (FL)

	IO VS FIX30 (matched)	ARM30 VS FIX30 (matched)	ARM30 & IO VS FIX30 (matched)
IO5	0.269***		0.185***
	(0.019)		(0.014)
	[1.309]		[1.203]
IO10	0.315***		0.215***
	(0.019)		(0.015)
	[1.370]		[1.239]
ARM30		0.164***	0.125***
		(0.017)	(0.015)
		[1.179]	[1.134]
Observations	51,886	62,388	92,234
Pseudo R-Squared	0.7949	0.8027	0.7893

Panel C Logistic Regression: State of Arizona (AZ)

	IO VS FIX30 (matched)	ARM30 VS FIX30 (matched)	ARM30 & IO VS FIX30 (matched)
IO5	0.361***		0.263
	(0.0248)		(0.021)
	[1.435]		[1.301]
IO10	0.445***		0.335
	(0.025)		(0.022)
	[1.561]		[1.398]
ARM30		0.206***	0.138
		(0.026)	(0.024)
		[1.229]	[1.147]
Observations	34,012	29,700	43,448
Pseudo R-Squared	0.7917	0.7807	0.7724

Panel D Logistic Regression: State of Nevada (NV)

	IO VS FIX30 (matched)	ARM30 VS FIX30 (matched)	ARM30 & IO VS FIX30 (matched)
IO5	0.326***		0.247***
	(0.035)		(0.031)
	[1.385]		[1.280]
IO10	0.386***		0.300***
	(0.034)		(0.032)
	[1.471]		[1.349]
ARM30		0.147***	0.111***
		(0.038)	(0.033)
		[1.159]	[1.117]
Observations	16,732	13,234	19,452
Pseudo R-Squared	0.8099	0.8129	0.8077

Note: This table shows results of logistic regression analysis for four representative states after Propensity Score matching, i.e., California, Florida, Nevada and Arizona. The dependant varaible is ‘default; takes the value of one for default loans and zero for others. We do not report the entire list of control variables, and refer to Table 2-3 for the full lists. The definitions of the independent variables are shown in Table 2. State, current year and origination year fixed effects are included in the regression but not reported. Standard errors are reported in parentheses, and estimated odds ratios for Logit regression are reported in brackets.
*Significant at the 10% level;
** Significant at the 5% level;
*** Significant at the 1% level.

Table 2-9 presents the sub-sample default analysis by setting the financial crisis as a break point in the matched sample. The results are consistent with those shown in Table 2-5: the default probabilities of both interest-only loans and 30-year adjustable-rate loans are higher than 30-year fixed-rate loans. However, there is a great deal of difference between the results before and after financial crisis. Before the financial crisis, borrowers of 5-year interest-only loans were around 20.4 percentage points more likely to default than those who selected 30-year fixed-rate loans, and borrowers of 10-year interest-only loans were around 7.8 percentage points more likely to default than those who selected 30-year fixed-rate loans. After the financial crisis, borrowers of 5-year interest-only loans were around 51.3 percentage points more likely to default than those who selected 30-year fixed-rate loans, and borrowers of 10-year interest-only loans are only around 47.5 percentage points more likely to default than those who selected 30-year fixed-rate loans. The results for 30-year adjustable-rate loans relative to 30-year fixed-rate loans are similar.

Table 2–9 Time Preferences and Mortgage Default: Financial Crisis Break Point of Propensity Score Matched Sample

	IO VS FIX_30	ARM_30 VS FIX_30	ARM_30 & IO VS FIX_30
Terminate before Financial Crisis			
IO5	0.186***		0.077***
	(0.018)		(0.012)
	[1.204]		[1.080]
IO10	0.075***		-0.027
	(0.020)		(0.015)
	[1.078]		[0.974]
ARM30		0.080***	0.062***
		(0.012)	(0.012)
		[1.083]	[1.064]
Observations	96,862	167,070	203,072
Pseudo R-Squared	0.1960	0.2664	0.2624
Terminate after Financial Crisis			
IO5	0.414***		0.296***
	(0.005)		(0.004)
	[1.513]		[1.345]
IO10	0.389***		0.303***
	(0.005)		(0.004)
	[1.475]		[1.354]
ARM30		0.210***	0.200***
		(0.005)	(0.004)
			[1.221]
Observations	512,398	492,250	789,000
Pseudo R-Squared	0.6962	0.6471	0.6603

Note: This table shows results of logistic regression analysis for financial crisis break point after Propensity Score matching. For simplify, here the time of financial crisis defined as the beginning of 2009. Therefore, the two parts of sample are: loans terminate before the end of 2008 and loans terminate after the beginning of 2009.The dependant variable is ‘default; takes the value of one for default loans and zero for others. We do not report the entire list of control variables, and refer to Table 2-3 for the full lists. The definitions of the independent variables are shown in Table 2-2. State, current year and origination year fixed effects are included in the regression but not reported. Standard errors are reported in parentheses, and estimated odds ratios for Logit regression are reported in brackets.
*Significant at the 10% level;
** Significant at the 5% level;
*** Significant at the 1% level.

2.6.3 Loans Originated between 2004 and 2007

The subprime mortgage market grew extremely quickly between 2001 and 2007. Kiff and Mills (2007), among others, argued that this was facilitated by the development of private-label mortgage backed securities, which do not carry the kind of credit risk protection offered by government-sponsored enterprises. Demyanyk and Hemert (2011) analysed loans that originated between 2001 and 2006 and found that, during the dramatic growth of the subprime (securitized) mortgage market, the quality of the market deteriorated dramatically. In addition, significant changes to Regulation C, which implemented the Home Mortgage Disclosure Act (HMDA), took effect in January 2004. These changes, designed primarily to enhance the understanding of mortgage markets and assist in fair lending enforcement, increased the amount and types of public information about residential real estate lending. Because of the dramatic growth of loans and new regulations, a sub-sample that consisted of mortgages originating between 2004 and 2007 is created and ran the same regressions as Table 2-3 and Table 2-6.[14] The results are consistent with what have been found in the previous sections (see Table 2-10).

[14] The sub-sample analysis of loans originated between 2004 and 2007 for four states and the financial crisis break point was also done, but the results were not reported in the chapter. The results are consistent with previous results.

Table 2–10 Time Preferences and Mortgage Default: Loans Originated between 2004 and 2007

Sample	Original Sample			Matched Sample		
Variables	IO VS FIX30	ARM30 VS FIX30	IO & ARM30 VS FIX30	IO VS FIX30	ARM30 VS FIX30	IO & ARM30 VS FIX30
IO5	0.350***		0.303***	0.371***		0.267***
	(0.003)		(0.003)	(0.005)		(0.004)
	[1.419]		[1.354]	[1.449]		[1.306]
IO10	0.389***		0.328***	0.358***		0.284***
	(0.003)		(0.003)	(0.005)		(0.004)
	[1.476]		[1.389]	[1.430]		[1.328]
ARM30		0.224***	0.196***		0.196***	0.172***
		(0.003)	(0.003)		(0.005)	(0.004)
		[1.268]	[1.217]		[1.216]	[1.188]
ownerocc	-0.185***	-0.215***	-0.188***	-0.240***	-0.232***	-0.249***
	(0.003)	(0.003)	(0.002)	(0.005)	(0.005)	(0.004)
	[0.831]	[0.810]	[0.829]	[0.786]	[0.793]	[0.780]
low_no_doc	0.326***	0.267***	0.312***	0.315***	0.255***	0.295***
	(0.003)	(0.003)	(0.002)	(0.005)	(0.005)	(0.004)
	[1.385]	[1.299]	[1.366]	[1.370]	[1.290]	[1.343]
subprime	-0.120***	-0.179***	-0.182***	-0.135***	-0.134***	-0.151***
	(0.004)	(0.004)	(0.003)	(0.006)	(0.007)	(0.006)
	[0.887]	[0.846]	[0.833]	[0.874]	[0.875]	[0.860]
OrigLTVRatioCalc	0.740***	0.641***	0.688***	0.733***	0.567***	0.729***
	(0.004)	(0.003)	(0.003)	(0.006)	(0.005)	(0.005)
	[2.096]	[1.892]	[1.989]	[2.081]	[1.762]	[2.072]
log_FICO Score	-1.019***	-0.995***	-1.048***	-0.984***	-0.859***	-1.038***
	(0.005)	(0.005)	(0.004)	(0.007)	(0.008)	(0.007)
	[0.361]	[0.372]	[0.351]	[0.374]	[0.423]	[0.354]

Continued

	Original Sample			Matched Sample		
Variables	ARM_30 VS FIX_30	IO & ARM_30 VS FIX_30	ARM_30 VS FIX_30	IO & ARM_30 VS FIX_30	ARM_30 VS FIX_30	IO & ARM_30 VS FIX_30
log _ Original Loan Balance	-0.206***	-0.159***	-0.175***	-0.154***	-0.134***	-0.184***
	(0.004)	(0.003)	(0.003)	(0.006)	(0.006)	(0.005)
	[0.814]	[0.823]	[0.840]	[0.858]	[0.875]	[0.832]
LOG_MHPI	-0.428***	-0.340***	-0.392***	-0.443***	-0.347***	-0.393***
	(0.004)	(0.004)	(0.003)	(0.007)	(0.006)	(0.006)
	[0.652]	[0.704]	[0.676]	[0.642]	[0.707]	[0.675]
log_duration	0.456***	0.749***	0.683***	0.414***	0.602***	0.585***
	(0.011)	(0.010)	(0.008)	(0.019)	(0.017)	(0.015)
	[1.578]	[2.674]	[1.980]	[1.512]	[1.825]	[1.795]
Observations	1,567,667	1,654,772	2,447,692	571,868	504,368	770,614
Pseudo R-Squared	0.6903	0.6460	0.6762	0.6831	0.6337	0.6510

Note: This table shows results of logistic regression analysis for BlackBox dataset with loans originated between 2004 and 2007: both before and after Propensity Score matching. The sample only includes 5-year interest-only loans, 10-year interest-only loans, 30-year fix-rate loans and 30-year adjustable-rate loans. The dependant varaible is ‘default; takes the value of one for default loans and zero for others. The definitions of the independent variables are shown in Table 2-2. State, current year and origination year fixed effects are included in the regression but not reported. Standard errors are reported in parentheses, and estimated odds ratios for Logit regression are reported in brackets.

*Significant at the 10% level;

** Significant at the 5% level;

*** Significant at the 1% level.

2.7 Conclusion

The financial crisis triggered by the subprime mortgage crisis meant that an economic recession spread from the US to the rest of the world. This led us to reconsider residential mortgage defaults and foreclosures. Mortgage default behavior is complex. Correctly identifying the different types of default behaviors that borrowers engage in is not only important for mortgage lenders and investors of mortgage backed securities, but also crucial for policy makers.

This chapter investigates whether heterogeneity in borrowers' time preferences as manifested in their mortgage choices correlates with their default decision. It presents evidence to explain the underlying theory of borrowers' unobserved heterogeneity and the source of this unobserved heterogeneity. Results indicate that present-biased borrowers who select back-loaded mortgages are more likely to default than dynamically consistent borrowers. In particular, naïve borrowers with present-biased preferences, who are more likely to select interest-only loans, default earlier than borrowers of other types of loans. If borrowers have present-biased preferences or suffer from short-term temptations and are aware of the consequences (termed as 'sophisticated' as opposed to 'naïve'), then it is likely they will prefer to refrain from the temptation and behave more rationally. The default of sophisticated borrowers follows the naïve borrowers. For dynamically-consistent borrowers, the choice of fixed-rate loans leads them to default less frequently than others. The relationship between present bias and mortgage default is maintained when controlling for borrowers' demographics and loan characteristics.

Overall, borrowers' heterogeneous time preferences as seen in their choice of mortgage types may help to better understand mortgage default behavior, and will assist in the creation of better policies to deal with the foreclosure crisis, such as mortgage modification and mortgage contract design.

Chapter 3 Reinforcement Learning and Mortgage Partial Prepayment Behavior

3.1 Introduction

In the mortgage market, prepayment and default risk are the two most important types of termination risks, and many studies have studied them and the corresponding behavior of borrowers (Kau et al. 1992; Kau, Keenan, and Kim 1994; Stanton 1995; Deng, Quigley, and Van Order 2000). Compared to full prepayment risk and default, little research has been conducted on the partial prepayment risk of borrowers in the residential mortgage market. Similar to default and full prepayment risk, partial prepayment also introduces risk to the duration of mortgage-backed securities (MBS) and affects their pricing. However, unlike default and full prepayment decision, the partial prepayment decision of a borrower only changes the unpaid mortgage balance. It does not terminate the mortgage contract, thereby allowing the borrower to perform repeated actions in the future and to learn from his/her early experiences when making future decisions.

The field of learning draws attention from various fields, such as economics, psychology, cognitive science, computer science, mathematics, and neural science (Bush and Mosteller 1955; Cross 1973; Arthur 1991, 1993; Roth and Erev 1995; Erev and Roth 1998). There has been growing interest in the effects of learning and earlier experience on an individual's decision-making (Ho and Chong 2003; Agarwal et al. 2006). The main stance taken in this literature is that agents react 'adaptively' when facing certain circumstances. For example, Erev and Roth (1998) found that reinforcement learning allowed for good predictions of an individual's future behavior. However, because of data-collection challenges, few papers have studied learning using individual micro-level data.

This chapter studies the risk of mortgage partial prepayment and the process through which mortgage borrowers learn to make partial prepayment decisions in the residential mortgage market in China. The learning dynamics of borrowers are measured by studying individual borrowers' repeated partial prepayments. Without terminating the mortgage contract, partial prepayments allow borrowers to learn from their early experiences, as well as those of others, to make repeated decisions in the future. In this chapter, reinforcement learning is manifested as the higher probability of borrowers who make more partial prepayments at earlier stages of their mortgage deciding to continue making partial prepayments, while controlling for other variables.

Since the reforms and the rapid development of the housing market from 1998, China's residential mortgage market has developed quickly. By the end of 2012, the total value of outstanding residential mortgages was 61 trillion RMB Yuan, approximately 9.9 trillion US dollars.[15] The mortgage environment in China is quite different from the US. US research on prepayment risk usually examines full prepayments. However, in China, partial prepayments are quite popular, because there is no penalty for prepayments. Thus, to study

[15] 1 Chinese Yuan = 0.162351 U.S. dollars

prepayment risk in China, one must distinguish between the probability of full prepayment and partial prepayment because a borrower can make these two decisions separately.

The empirical analysis in this chapter uses a rich set of individual mortgage payment history data from a leading mortgage lender in China. The informative loan history dataset contains not only mortgage loan information, but also borrowers' characteristics and their payment decisions for their mortgages in each period.[16] This longitudinal dataset contains information on 172,328 individual loans that originated between 2003 and 2010 with 5,282,182 monthly payment events.[17] This chapter focuses on partial prepayments, and asks whether the probability of mortgage borrowers choosing to partially prepay is higher once they have gained more experience. The empirical model is based on the conditional fixed effects multinomial logit model (FEMNL) (Rasch 1960; Chamberlain 1980).

The results indicate that 'option theory' does not play a significant role in determining mortgage partial prepayments in China. Instead, other financial factors, such as stock market investment opportunities, play a major role. Secondly, borrowers' characteristics, such as age, job position, gender, and income, are important indicators for predicting borrowers' partial prepayment behavior. In addition, following Deng and Quigley (2012), a behavioral

[16] At each period, borrowers can choose from: 'default', 'paid off/ full prepayment', 'partial prepayment', and 'continue make monthly payment'. In China, the minimum amount of partial prepayment is 10,000 Chinese Yuan, and there is no prepayment penalty for both full prepayment and partial prepayment.

[17] The original database contains a large number of mortgage observations and payment events on single family mortgage loans issued from 2003 to 2010. The regression data used in this chapter is based upon a random sample of ten percent of those mortgages loans.

correlate of the unobserved heterogeneity of individual borrowers is created and added into the model to improve the estimation of mortgage holders' responses. Lastly, and most importantly, a borrower's partial prepayment behavior follows the reinforcement learning process. A borrower's partial prepayment decision depends not only upon current stage variables (such as other investment opportunities that he/she can engage in) and his/her characteristics, but also past experiences. Borrowers can learn from their own experiences, as well as those of others. Borrowers who make more partial prepayments early on have a higher probability of making the same decision in the future. In particular, the self-learning experience increases the probability of partial prepayment by around 26.9 percentage points, and the experience of learning from others increases the probability of partial prepayment by around 1.8 percentage points.

Moreover, the results show that learning dynamics are not monotonic, as borrowers act as if their knowledge depreciates – i.e., learning patterns exhibit a recency effect. Recent experience plays a larger role than older experience in determining the partial prepayment behavior of borrowers.

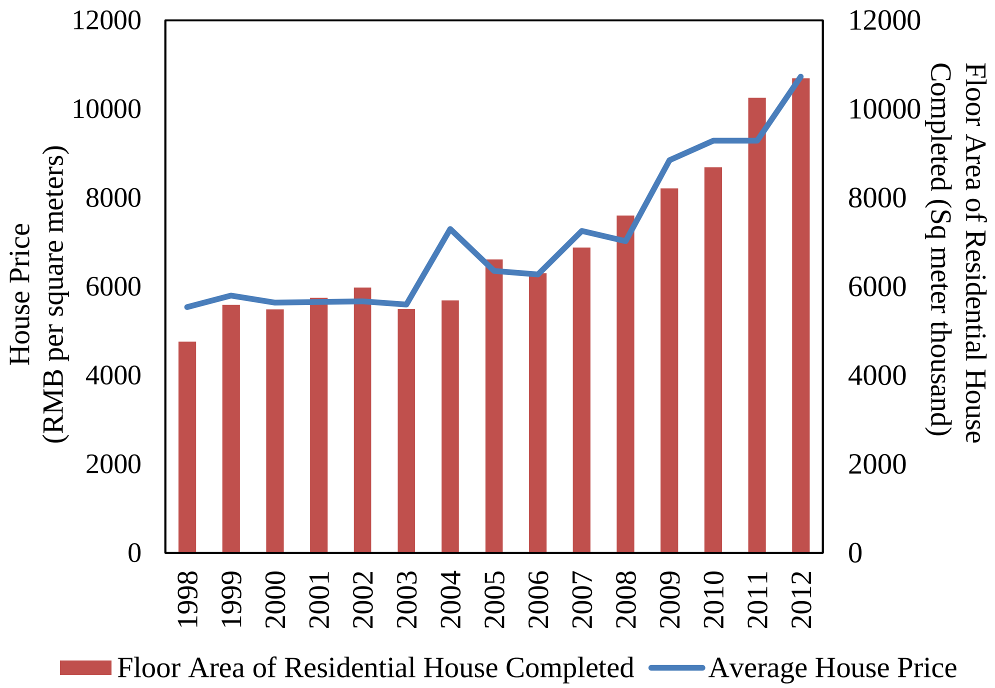
This chapter proceeds as follows: in section 3.2, a brief introduction of house prices and the mortgage market in China is given; section 3.3 presents the relationship between this chapter and prior literature; Section 3.4 summarizes the data and presents the empirical evidence for learning and backsliding; the last section concludes this chapter.

3.2 The House Price and Mortgage Market in China

China is the largest developing economy and its housing market has increasingly attracted academic attention. Since the founding of the People's Republic of China (PRC) in 1949, the housing market in China has experienced several waves of reform. A milestone reform event happened in 1998 with the issue of the 23rd Decree: housing was no longer allocated to citizens, kick-starting the modern private housing market. From then on, the government would no longer distribute housing to the public and all households were required to buy or rent a house from the private housing market. This change brought about a new stage of development in the Chinese housing market. The number of privately-built houses and house prices began to grow dramatically. According to the National Bureau of Statistics of China, investment in China's real-estate sector was 30 trillion Chinese Yuan (4.5 trillion US Dollar) in 2008, having increased by 20.9% compared to the previous year.

Figure 3-1 shows the average house price and floor area of residential houses that built from 1998 to 2012. It can be seen that both house prices and the number of houses built increased remarkably in this period. House prices increased by about 193% over the 15-year period from 1998 to 2012, while the number of completed residences increased by more than two times in the same period. House prices increased from 2004, mainly because of the launch of a public land auction and listing system, with the first land auction in China being held in Shenzhen in 1987. However, from 1987 to 2004, there were no public auctions of land parcels. Developers were required to contact local governments about land parcels they were interested in, and they would then negotiate a price without an auction. From 2004, a policy that all residential and commercial urban land had to be listed and auctioned publicly was implemented (Wu, Gyourko, and Deng 2012). From then on, all developers were required to bid at auctions for the land they desired, which may have contributed to the obvious increase in house prices from 2004.

Figure 3-1 House Price and Floor Area of Residential House Completed in China

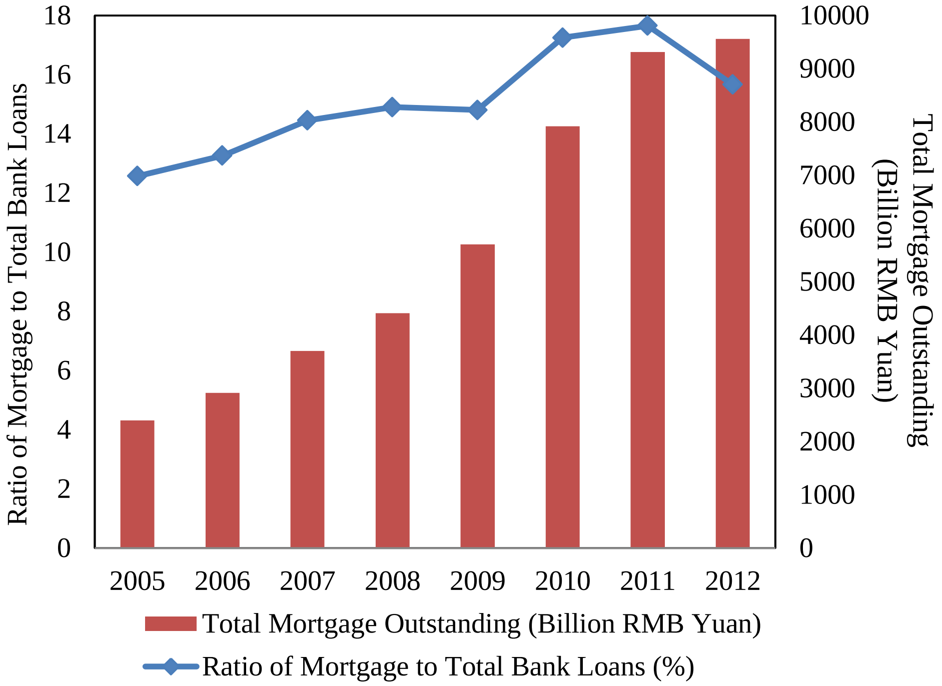


Note: This is from National Bureau of Statistics. The house price here is the average house price across the whole country, and is calculated as 'Total Residential House sale'/'Total Floor Area of Sale'. The floor area of completed residential houses is the total floor area that has been completely built. The left Y-axis measures house prices in Chinese Yuan per square meters, while the right Y-axis measures the floor area of completed residential houses in thousand square meters. The X-axis indicates the year.

The rise of the booming real estate market also contributed to the development of the mortgage market in China. From 2005 to 2012, the outstanding balance of mortgage loans increased nearly four-fold. According to the "Statistical

Report of Loans of Financial Institutions in 2012" from the People's Bank of China, up to the end of 2012, commercial banks held nearly 9.5 trillion Chinese Yuan worth of residential mortgages and the share of residential mortgages in the total value of loans they made rose to 16% from 4% between the years 2005 to 2012 (Figure 3-2).

Figure 3-2 Residential Mortgage Market in China



Note: This is from The People's Bank of China. The left Y-axis measures the ratio of mortgages to total bank loans, while the right Y-axis measures the total mortgage outstanding in billion Chinese Yuan. The X-axis indicates the year.

Four commercial banks mainly issue residential mortgage loans in China: Industrial and Commercial Bank of China (ICBC), China Construction Bank (CCB), Bank of China (BOC) and Agricultural Bank of China (ABC). China Construction Bank (CCB) was the first bank to issue residential mortgage loans in China. Several distinctive features of China residential mortgage loans are introduced in the following sections.

3.2.1 Borrowing Requirements

Borrowers should have a stable source of income and a good credit record, and be between 18 to 65 years of age. Generally, the loan-to-value ratio should be lower than 30%, and the term of the loan should be less than 30 years. To apply for a mortgage loan, applicants should provide a real estate certificate or purchase contract and proof of down payment from the developer, proof of income (this is the main document for housing mortgage loan applications in China), evidence of other types of property (such as another real estate certificate, stocks, funds, cash deposits, vehicle permits, etc.). According to the requirements announced by the China Banking Regulatory Commission (CBRC) in the 2004 "*Guidelines for the risk management of real estate loans of commercial banks*", the ratio of monthly mortgage payments to income for borrowers should be lower than 50%, and the ratio of total monthly debt payment to income should be lower than 55%.

3.2.2 The Scope of Collateral

In China, the collateral for mortgages can only be houses. This includes villas, with the down-payment ratio for a villa being higher than that for other types of houses. The age of the house (from the housing completion date) usually should be no more than 20 to 30 years, and the sum of the age of the house and the loan term should be no more than 30 to 40 years. In other countries, such as US, the collateral of a mortgage can belong to the borrower or others. If the collateral belongs to others, the mortgagor must get the permission and signature of the property owner and their spouses. However, in China, the collateral of a mortgage can only belong to the borrower himself.

3.2.3 Loan Application Procedure

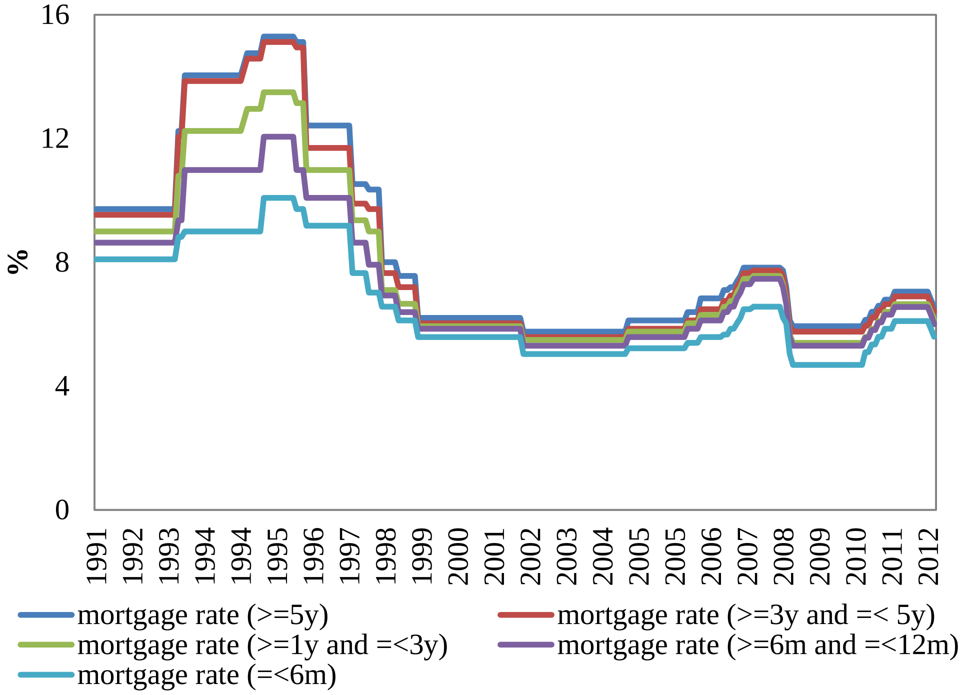
Applicants should first submit the required documents. After receiving the application form filled by the applicant together with the relevant documents, the bank carries out eligibility investigations. The most important factor for the bank to investigate is the income statement and credit record. Individual credit records can be checked by the rating system of the People's Bank of China. This rating system was on trial in December 2004, and began running officially from January 2006. Upon approval, the bank and the borrower sign a mortgage contract. The borrower then opens a mortgage account at the mortgage bank for making mortgage payments. Every month, the borrower makes a specified payment to the bank according to the mortgage contract.

3.2.4 Mortgage Interest Rate

In China, the mortgage interest rate is regulated by the central bank of China, known as the People’s Bank of China. Interest rates determined by the People's Bank of China can be executed by commercial banks after approval by the State Council. All banks are expected to follow the lending rules set by the People’s Bank of China. According to the People's Bank of China’s regulations, the mortgage interest rate should be a multiple of the benchmark lending rate. Before August 2006, the multiple was 0.9; from August 2006 to October 2008, the multiple was 0.85; from October 2008 to March 2010, the multiple was 0.7; after March 2010, the multiple was 0.85 for the first house, and 1.1 for the second house. In addition, before 2010, the real mortgage interest rate was the lowest interest rate regulated by the People's Bank of China. However, since 2011, because of the reach of credit risk measurement techniques and other tools of banks, mortgage interest rates higher than the lowest interest rates regulated by the People's Bank of China started to increase. Currently, the benchmark lending interest rate is 5.6% for mortgage loans with a term of 6 months or below. For loans with a term above 5 years, the benchmark lending interest rate is 6.55%. The spread between long term and short term is 95 basis points.

One special feature of Chinese mortgage loans is that all loans in the current market are adjustable rate mortgages (ARM), and there are no fixed rate mortgages (FIX). Some commercial banks issued a few fixed rate mortgages during 2007 to 2008, but these disappeared in a very short time. During the mortgage payment term, if the People’s Bank of China changes the interest rate, the interest rates of all mortgage loans will be adjusted according to the new interest rate. A few mortgage loans are adjusted in the next month or next quarter, while the majority of mortgage loans will be adjusted on the first day of the next year. Figure 3-3 shows the lending interest rates announced by the People’s Bank of China from 1991 to 2012.

Figure 3-3 Lending Rates in China



Note: This is from The People's Bank of China. In China, mortgage interest rates are determined by the People’s Bank of China, the central bank of China. All banks follow the lending rules that it sets. *6m*: six months; *12m*: twelve months or one year; *1y*: one year; *3y*: three years; *5y*: five years.

3.2.5 Payment Method

If the loan term is one year or less, both principal and interest must be repaid as a lump sum at maturity. If the loan term is greater than a year, the loan may be repaid in equal instalments of the principal plus interest, or in equal instalments of the principal. The borrower may choose either method, but there is only one payment method for each loan, and after the method has been specified in the contract, it may not be changed. Loan applications state that once a mortgage contract has been signed, borrowers should open a mortgage account at the mortgage bank for making mortgage payments. Borrowers should make a specified monthly payment to their bank according to the mortgage contract. Each month, borrowers can choose different payment decisions on their mortgage: continue paying, default on their payments, make a full prepayment, or make a partial prepayment.

3.2.6 Mortgage Termination

Default and full prepayment are two channels for terminating a mortgage contact during the mortgage's term. There are very few defaults in China. Besides the potential cultural reasons, one main reason for this is that, unlike the United States, all mortgage loans in China are recourse loans. This allows mortgage lenders or banks to recover their loan losses from the borrowers' assets. Thus, once a borrower defaults, all of his/her assets will be taken away to cover the loss of the mortgage lenders.

Full prepayment and partial prepayment are very popular in China. The motivation for making a full prepayment in China is quite different from doing so in the United States because of the unique way in which mortgage interest rates are set: once the People's Bank of China announces a rate change, all mortgage interest will be adjusted according to the change, and all banks use the same lending rate benchmark. Hence, all prepayments observed in the sample are payoffs or partial prepayments rather than refinances. In China, mortgage refinance is not allowed.[18]

[18] Refinancing is the process of paying off an existing loan by taking a new loan and using the same property as security. This is not allowed in China.

3.3 Relationship to the Prior Literature

This chapter adds to several strands of the existing literature. First, there is a large volume of literature on the risk of mortgage lending and the termination behavior of borrowers in the United States. Pioneered by Asay (1978), there was a quick expansion of studies on mortgage valuation and borrower behavior based on contingent claims models, mainly developed by Black and Scholes (1973), Merton (1973a), and Cox, Ingersoll, and Ross (1985). The contingent claims model provides a useful framework for analyzing borrowers' termination behavior: prepayments are treated as an American call option and default as a compound put option. Most studies use option models to explain borrowers' termination behavior, whether they do so using a full prepayment or by defaulting on their payments, or by doing both. There are a few studies on partial prepayment behavior. For instance, Dunn and McConnell (1981b) modelled the optimal full prepayment strategy of a mortgage holder, where full prepayment was regarded as a call option. Buser and Hendershott (1984), and Brennan and Schwartz (1985) also used option models to price the risk of full prepayment. Schwartz and Torous (1989) empirically modelled prepayment as a function of exogenous or explanatory variables in a regression model.

Some other researchers use option theory to explain default behavior. Cunningham and Hendershott (1984) used the option approach to derive mortgage default insurance premiums using a sample of FHA loans, with mortgage default treated as a put option. Titman and Torous (1989), and Kau, Keenan, and Kim (1994) applied option models to mortgage defaults, and concluded that well-informed borrowers will default immediately when the mortgage value exceeds the property value at any time during the loan term. However, Foster and Van Order (1984) noted that borrowers would not default "ruthlessly", and exercised the put option of default if the value of house fell below the mortgage value by an amount equal to the net transaction costs. Therefore, the argument that transaction costs matter in mortgage default is important (Cunningham and Hendershott 1984; Foster and Van Order 1984, 1985; Vandell and Thibodeau 1985; Quigley and Van Order 1991; Lekkas, Quigley, and Van Order 1993).

A series of papers provide support to emphasize the importance of the relationship between full prepayment and default options (Titman and Torous 1989; Kau et al. 1992). Deng, Quigley, and Van Order (1996) and Deng (1997) were the first to model residential mortgage prepayments and defaults as a joint decision using micro-level data. Deng et al. (2000) modelled residential mortgage prepayments and defaults as competing risks, and considered the issue of unobserved heterogeneity in the context of hazard modelling. However, few studies pay attention to partial prepayment risk and the corresponding behavior of borrowers.

In addition, there is limited literature on China's mortgage markets, as it has been impeded by data limitations. Deng, Zheng, and Ling (2005) was the first rigorous empirical work to study the risk of residential mortgage markets in China, noting that while the option theory failed to explain prepayment and default behavior in the residential mortgage market in China, the behavior could be explained by other financial factors that were unrelated to option theory. The same authors also found that borrower characteristics were significant in determining borrowers' prepayment behavior, and may thus be used as an effective tool for screening potential high-risk borrowers in the loan origination process. Deng and Liu (2009) studied the termination risk of mortgages in the Chinese housing market by using embedded forward contracts. They found that borrower characteristics and collateral information are both important in determining mortgage termination risks in China. Overall, very few papers study the Chinese mortgage markets, and this chapter addresses this gap to improve our understanding of Chinese mortgage markets.

This chapter also adds to literature on reinforcement learning, which has been extensively addressed both in psychology and behavioral economics. In psychology, Bush and Mosteller (1955) proposed the first mathematical model of reinforcement learning. The Bush-Mosteller model was later adapted and generalized in economics by Cross (1973), Arthur (1991, 1993), Roth and Erev (1995), and Erev and Roth (1998). However, many economic studies have analyzed learning in laboratory environments. Due to data limitations, only a few studies have measured learning with household-level data. For example, Miravete (2003) and Agarwal et al. (2006) respectively showed that consumers switch telephone calling plans and credit card contracts to minimize monthly bill payments. Other researchers have shown the predictive power of learning models. For example, Ho and Chong (2003) used grocery store scanner data to estimate a model in which consumers accumulate not only product-level experience but also attribute-level experience, and they learn from these experiences to make decisions. Agarwal et al. (2006) studied the learning process in credit card market, and indicated that the speed of net learning was about twice as great for higher-income borrowers than it was for lower income borrowers. At the same time, the rate of knowledge depreciation, or forgetting, was about half as fast for high-relative to low-income borrowers. Middle-aged borrowers have the same advantageous learning dynamics relative to older borrowers. Haselhuhn et al. (2012) studied video stores, and found that renters were more likely to return their videos on time if they had recently been fined for returning them late. In this chapter, individual loan-level data is used to test for the existence of learning in the mortgage market. This study will provide evidence for the role of learning in explaining the partial prepayment behavior in the mortgage market.

3.4 The Empirical Analysis

3.4.1 Empirical Methodology

Following Axel (1990), in the empirical analysis of this chapter, a conditional fixed effects multinomial logit model (FEMNL) is employed to examine the partial prepayment behavior of mortgage borrowers and their learning process. In the FEMNL model, the choice probabilities of borrowers are conditional on their past learning experiences and have the same convenient multinomial logit form as their unconditional choice probabilities.

Let $P_{nt}(.)$ denote the probability that household n, n=1, …, N, makes decisions on its mortgage payments: 'd' (default), 'p' (full prepayment), 'c' (continue to pay) and 'i' (partial prepayment), in period t, t=1, …, T. The choice of each payment decision is determined by a vector of explanatory variables which vary by each borrower, time, and loan, X_{int}, and a learning factor for each borrower s_n. The learning factor s_n consists of a self-learning factor and a factor to measure learning from others, where:

(3.1) $$s_{in} = \sum_{t=1,2,\dots,T} d_{int}$$

$$d_{int} \begin{cases} = 1 \ if\ borrower\ n\ choose\ i\ at\ time\ t \\ = 0 \qquad\qquad\qquad\qquad otherwise \end{cases}$$

For self-learning, s_{in} is a sufficient statistic and can be observed from the data. In addition to self-learning, borrowers can also learn from others. The calculation of the factor of learning from others is very similar to the self-learning factor.[19]

Algebraically, conditioning on the s_{in}, yields the following conditional choice probabilities for the *nth* borrower's sequence of choices over time, denoted by $y_n = (y_{n1}, \dots, y_{nT})$ (Chamberlain 1980):

(3.2) $$\text{prob}(y_n | s_n)$$

$$= \exp\left(\sum_{t=1,\dots,T} \exp(X_{int}(I_B, I_L, I_F)\beta) \middle| j = y_{nt} \right)$$

$$/ \sum_{D \in B_n} exp\left(\sum_{k=d,p,c,i;\ t=1,\dots,T} D_{kt}\, X_{knt}(I_B, I_L, I_F)\beta \right)$$

[19] Learning from others is shown here, which is similar in form to learning from self, where:

$$s_{in} = \sum_{i=4} d_{i(-n)t}$$

$$d_{int} \begin{cases} = 1 \ if\ borrower(-n) choose\ i\ at\ time\ t \\ = 0 \qquad\qquad\qquad\qquad otherwise \end{cases}$$

Borrower (-n) represents all other borrowers around borrower n, and borrower n can learn from them (-n). In the empirical work, borrower (-n) and borrower (n) are from the same company.

Where the borrower and time varying explanatory variables, X_{int} mainly include three groups information: borrowers' characteristics I_B , loan information I_L, and intrinsic values of the default and prepayment options I_F.

The set B_n represents the set of all choice sequences that result in the same "aggregate choice pattern" $s_n = (s_{dn}, s_{pn}, s_{cn}, s_{in})$ of the household:

(3.3) $$B_n = \left\{ D = (D_{kt})_{k=d,p,c,i;\, t=1,\dots,T} \middle| D_{it} \in (0,1), \sum_k D_{kt} = 1, \sum_t D_{kt} = s_{kn} \right\}$$

The choice probabilities in Equation (3.2) are multinomial logit choice probabilities conditional on choice set B_n that varies by household. The resulting conditional loglikelihood function is:

(3.4) $$\mathrm{L} = \sum_{n \in C} \log prob(y_n | s_n)$$

It needs to be accumulated only over the set C that includes all borrowers that have chosen mortgage partial prepayment at least once:

(3.5) $$\mathrm{C} = \{n = 1, \dots, N | max_i(s_{in}) \leq T - 1\}$$

and can be estimated with conventional logit packages.

The coefficients β in the FEMNL model have the same interpretation as the in the conventional logit model and can be used to calculate choice elasticities with respect to the *k*th explanatory variable:

(3.6) $$E_{ij}^k = d \log P(i) \ / d \log X_j^k = \beta^k \cdot X_j^k \cdot \left(D_{ij} - P(j)\right),$$

where the omission of the borrower index n and the time index t refers to the mean across all borrowers and all time periods, and $D_{ij} = 1$ if and only if i=j.

3.4.2 Data Collection

The empirical analysis is based on a unique individual mortgage dataset with loan history information collected by a major residential mortgage lender in China. The dataset contains a large number of mortgage observations and monthly payment events on single family mortgage loans in 35 major cities in China issued from 2003 to 2010. All loans are adjustable rate mortgage loans. For each loan, the available information includes static information taken at the time of origination, such as the mortgage date, the original loan amount, the initial loan-to-value ratio, the mortgage contract interest rate, term, and the province and city in which the property is located. The data also include dynamic data on monthly payments, mortgage balances, and indictor of full prepayment, partial prepayment, and default. Besides loan information, the dataset also provides valuable information about borrowers' characteristics, including monthly income, age, gender, marriage status, education, occupation, job position and houses and mortgages they currently have. The regression data used in this chapter is based upon 172,328 mortgages loans with 5,282,182 monthly payment events.

3.4.3 Regression Variables

Table 3-1 provides a concise summary of all the regression variables used in the empirical analysis. Some key variables and their derivations are discussed below.

Table 3–1 Regression Variables Description

Variables	Definition and Description
Woodheads Factor	Following Deng and Quigly (2012), the woodheads factor in this chapter reflects difference in "astuteness' among borrowers' investment choice. This factor can be viewed as a correlate of the unobserved heterogeneity of individual borrowers.
Self-Learning Factor	This reflects the learning borrowers gain from their own experiences. It can be calculated from Equation 3.1.
Learning from Others	This reflects the learning borrowers' gain from the experiences of others. It can be calculated from footnote19.
Option Variables:	
Call option	Call options are different in China compared to the US. This difference is an outcome of the probability of negative interest rate, and reflects investment opportunity costs in China.
Put option	Probability of negative equity at termination or censored.
No. of House and Mortgage Loan Possessed:	
2 houses and 1 mortgage	Borrower with 2 houses and 1 mortgage (1=Y).
2 houses and 2 mortgage	Borrower with 2 houses and 2 mortgage (1=Y).
Education:	
Graduate and above	Borrower's education level is graduate and above (1=Y).
Income:	
Log monthly income	The logarithm of borrower's monthly income
With Other Income more than 10,000	Whether the borrower has other annual income more than 10,000 Chinese Yuan (1=Y), in addition to his/her regular monthly income.
Age>=40	Borrower is older than 40 (1=Y)
First house	This is the first house that borrowers have brought and are using as collateral (1=Y).
Married	Borrower's marriage status (1=Y)
With Zhi Cheng	'Zhi Cheng' refers to the level of professional's ability and professional achievement. In China, Zhi Cheng includes five levels: Senior, Associate senior, intermediate, assistant, technician level (1=Y).
Male	Borrower is male (1=Y)
Dependent	Dependent number of borrower
LTV	Loan to value ratio
Log loan amount	The logarithm of loan amount

Note: This table presents the regression variables in this chapter. Besides loan and borrower characteristics, the most important variables in this chapter are 'Self-Learning Factor' and 'Factor of Learn from Others', and their definitions can be found in the following discussion.

The variables '*Number of House and Mortgage loans possessed*' indicates the extent of payment pressure that borrowers face. The term '*Borrower with 2 houses and 1 mortgage*' refers to those who have two houses, where one is financed with a mortgage and the other is fully owned. They can rent out one of their houses to relieve their mortgage pressure. Borrowers with 2 houses and 1 mortgage face a lower level of payment pressure than those with 1 house and 1 mortgage. The level of payment pressure for borrowers with 2 houses and 2 mortgages should be lower than the payment pressure for those with 1 house and 1 mortgage, but higher than those with 2 houses and 1 mortgage.

a) Option variables

Since the early works of Dunn and McConnell (1981) and Green and Shoven (1986), researchers have modelled mortgage contracts in a contingent claims framework: a borrower's option to prepay the mortgage is an embedded call option at a strike price of par while the default option is a put option at a strike price equal to the market value of the collateral property. In the United States, there are two primary motivations for borrowers to exercise their call option: (1) to refinance their existing debt at a lower rate of interest; or (2) to terminate their debt through sale of the underlying asset.[20] If the current market value of the house, which serves as collateral of the mortgage debt, drops below the current value of the remaining mortgage balance, a borrower has an incentive to default. In the absence of transaction costs, a rational borrower can maximize his/her welfare by exercising the options when they are "in the money".

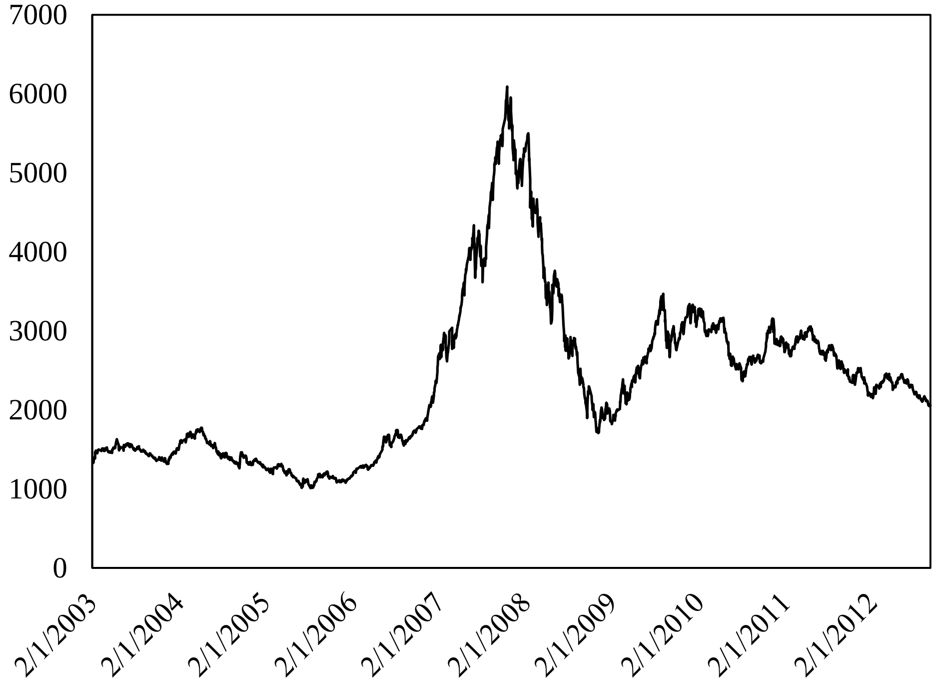
---

[20] Borrowers may sell their house for various reasons such as relocation for work and changes in family circumstances.

As discussed above, the special features of Chinese mortgage contracts have to be taken into account when calculating the option values. In China, the mortgage interest rate is regulated by the People’s Bank of China. All changes in this rate are subject to changes in the lending rates published by the People’s Bank of China. In addition, refinancing is not allowed in China. Based on these considerations, the value of the call option is calculated differently in China than the US. In this chapter, the calculation of the call option follows Deng and Liu (2008), where the prepayment option does not depend on the interest rate, but is instead closely related to the borrower's alternative investment set.

Mortgage debt is treated as a consumption smoothing instrument. There are very few investment opportunities and very few borrowing vehicles in the current Chinese capital market. The mortgage market is a major and steadily growing sector in the Chinese debt market while the stock market is the major investment sector. Therefore, borrowers will make prepayment decisions based on the cost of capital and stock market returns. They will prepay when the cost of capital (mortgage rate) exceeds the investment returns. Figure 3-4 shows the stock index return in China from 2003 to 2012. The optimal stopping time of prepayment depends on a borrower's income and his/her judgment about stock market returns and interest rates in the future. If borrowers are in the same circumstances, i.e. they have same income flow, same information, same perceptions of the macro economy, and the same level of risk aversion, they will make prepayments.

Figure 3-4 Stock Index in Shanghai Exchange of China



Note: The Shanghai Stock Exchange Composite Index is a capitalization-weighted index. The index tracks the daily price performance of all A-shares and B-shares listed on the Shanghai Stock Exchange. The index was developed on December 19, 1990 with a base value of 100. Index trade volume on Q is scaled down by a factor of 1000.

The "put option" in this chapter is measured by the probability of negative equity, and the calculation follows Deng, Quigley and Van Order (2000). The put option can be calculated as:

$$\text{(3.7)} \qquad \text{PUT OPTION} = \emptyset\left(\frac{\log\left(V^{*}_{i,r_i}\right) - \log\left(M_{i,k_i}\right)}{\sqrt{\sigma_t^2}}\right)$$

where $\Phi(.)$ is the standard normal cumulative distribution function, V^{*}_{i,r_i} is the current market value of mortgage debt and M_{i,k_i} is the current market value of the property i. σ_t is housing index volatility. The current market value of mortgage debt is calculated using the current interest rate and monthly payments which are obtained from the database:

(3.8) $$\mathrm{V}^{*}_{\mathrm{i},\mathrm{r}_\mathrm{i}} = \sum_{\mathrm{i}=1}^{\mathrm{T}-\mathrm{k}} \frac{\mathrm{M}_{\tau+\mathrm{k}}}{(1+\mathrm{r}_{\tau+\mathrm{k}})^{i}}$$

where M is monthly payment of mortgage principle and interest, $\mathrm{r}_{\tau+\mathrm{k}}$ is adjustable mortgage rate for loan originated at time τ and after the seasoning period of k. T is the mortgage term.

The current market value of property is calculated as:

(3.9) $$\mathrm{M}_{\mathrm{i},\mathrm{k}_\mathrm{i}} = \frac{\mathrm{ORIBAL}}{\mathrm{LTV}} \times \frac{\mathrm{H}_\mathrm{t}}{\mathrm{H}_0}$$

with LTV is the original loan-to-value ratio which is indicated in the database, H_t is the housing price index.

The intrinsic value of the call option is:

(3.10) $$\mathrm{CALL\ OPTION} = \frac{\mathrm{V}^{*}_{\mathrm{i},\mathrm{r}_\mathrm{i}} - V_R}{\mathrm{V}^{*}_{\mathrm{i},\mathrm{r}_\mathrm{i}}}$$

where $\mathrm{V}^{*}_{\mathrm{i},\mathrm{r}_\mathrm{i}}$ is defined the same way as in the "put option", as the current market value of the mortgage, that is, the cost of financing a house purchase, and V_R is the value of a hypothetical income, that is, the return from an alternative investment. Since the stock market is a major investment alternative, the Shanghai stock price index is used in calculating the return of the investment. Specifically, V_R is defined as:

(3.11) $$V_R = \sum_{i=1}^{T-k} \frac{M_{\tau+k}(1+R_{\tau+k})^{i}}{(1+r_f)^{i}}$$

where $M_{\tau+k}$ is, as previously stated, is monthly payment of mortgage principle and interest. r_f is the risk free interest rate. In this chapter, the risk free rate is represented by the basic lending rate from the People's Bank of China.

b) Self-Learning Factor

The calculation for the self-learning factor can be found in Equation 3.1. In the data, it can be observed that the number of times borrowers have partially prepaid their mortgages before the current decision time each month since origination. Then, the self-learning factor for each borrower at each time period can be computed, reflecting how often he/she has chosen to make a partial prepayment in the past. It is a time-varying variable. Self-learning is expected to have an effect on the likelihood of partial prepayment: the probability that a borrower will make a partial prepayment in the future positively depends on the number of partial prepayments he/she has made previously. Therefore, borrowers who have made more partial prepayments in the earlier stages of their mortgage will be more likely to continue making partial prepayments in the future.

c) Learning From Others

The calculation of the impact of learning from others is specified in footnote 19. The dataset records the company identification numbers of the workplaces of the borrowers. An assumption is made that borrowers from the same company will interact with and learn from each other. For each month since the origination of a loan, the number of times all other borrowers from the same company made a partial prepayment on their mortgages in previous time periods before the current decision time can be observed. Therefore, the impact of learning from others is computed as the total number of times that all other borrowers from the same company made partial prepayments before the current time period. Just like the self-learning factor, it is a time-varying variable. The expected effect of learning from others on the partial prepayment decision is: the probability that borrowers will make partial prepayments in the future is positively related to the earlier partial prepayment experiences of their peers in the firm where they work. Thus, the probability that borrower *A* will make a partial prepayment is higher if more people from his/her workplace have been making partially prepayments.

d) Woodheads Factor

In the mortgage market, some correlates of unobserved heterogeneity of individual borrowers are observed in the data, and a woodheads factor was created to reflect differences in “astuteness’ among borrowers (see Deng and Quigly 2012). The woodheads factor in Deng and Quigly (2012) is similar to the burnout effect, which reflects how pools of mortgage loans which have experienced large exposure to refinancing opportunities tend to have lower prepayment rates, other things being equal. Each month since origination, the call option’s status (that is, whether it is in the money) is calculated. Then the woodheads factor for each borrower is computed: it reflects the number of months since origination that an in-the-money call option was not exercised by partial prepayment. However, the woodheads factor in this essay is different from that of Deng and Quigly (2012). The calculation of the value of the call option of borrowers in this essay is based on the cost of capital and the stock market return. In a perfect market, borrowers should choose investments with a higher return rate. Thus, the woodheads factor in this essay reflects difference in “astuteness” among borrowers’ investment choices, instead of refinancing opportunities.

3.4.4 Descriptive Statistics

Table 3-2 shows the basic statistics for the variables in the empirical model by loan characteristics (origination year, loan to value ratio, loan amount) and household characteristics (such as marriage status, gender, age, house owned and mortgage loan, occupation, education and “zhicheng”[21]).

[21] The definition of “Zhicheng” is: the title of a technical or professional post (such as engineer, professor, lecturer, academician, etc.). It reflects their technological capability and work capacity.

Table 3–2 Descriptive Statistics for Mortgage Loans
-Frequency of Loans by Major Categorical Covariates

Variables	Loan Observations	Percentage on Total Loans
Origination Year		
2003	9,818	5.7
2004	16,330	9.48
2005	13,939	8.09
2006	16,104	9.34
2007	27,195	15.78
2008	18,683	10.84
2009	46,951	27.25
2010	23,308	13.53
Loan to Value Ratio		
LTV<=50	44,114	25.60
50<LTV<=60	24,784	14.38
60<LTV<=70	64,193	37.25
70<LTV<=80	39,227	22.76
Original Loan Amount (RMB)		
OLA=<200,000	59,528	34.54
200,000 <OLA<400,000	60,867	35.32
OLA >=400000	51,933	30.14
Marital Status		
Married	103,660	60.15
Single	68,668	39.85
Gender		
Male	84,617	49.10
Female	87,711	50.90
No. of House and Mortgage Loan Possessed		
Borrower with2 houses and 1 mortgage	53,813	31.23
Borrower with1 house and 1 mortgage	18,784	10.90
2 houses and 2 mortgages	5,926	3.44
Occupation		
Official	2,583	1.50
Institution	9,921	5.76
Professional	5,454	3.16
Education		
Graduate and above	51,601	29.94
Others	120,727	70.06
Age		
Age>=40	34,806	20.20
Age<40	137,522	79.80
Zhi Cheng		
With Zhicheng	26,281	15.25
Without Zhicheng	146,047	84.75
Total Observations	172,328	100

Note: This table shows the basic statistics of regression variables in the empirical model by loan information and household characteristics. The definitions of the variables are shown in Table 3-1.

Table 3-3 shows the statistics for the woodheads factor, M, for mortgage loans and mortgage payment events. Panel A shows the distribution of M by mortgage loans, separately for the full sample and for differently seasoned mortgage pools.[22] It can be seen that nearly 91% of the mortgage loans in the sample have missed at least one opportunity to invest in stock markets to earn a higher return rate. About 55.71% of borrowers in the two years' seasoned pools have missed more than twelve opportunities, while for five years' seasoned pools, the percentage is smaller at 44.06%. The results for the payment events listed in the Panel B are calculated similarly to Panel A. It shows the distribution of M by payment events, separately for the full sample and for differently seasoned mortgage pools. Nearly 90% of payment events in the sample missed at least one opportunity to change their investment as stock.

[22] The two year seasoned pool is a sub-sample of mortgage loans whose durations are greater than two years. The three and five years seasoned pools have a similar intuition, meaning the sub-sample of mortgage loans with durations greater than five years or ten years. As indicated in Deng and Quigley (2012), the full sample may be interpreted as a pool containing the newly issued mortgage loans, like duration year bigger than 0 but smaller than 3.

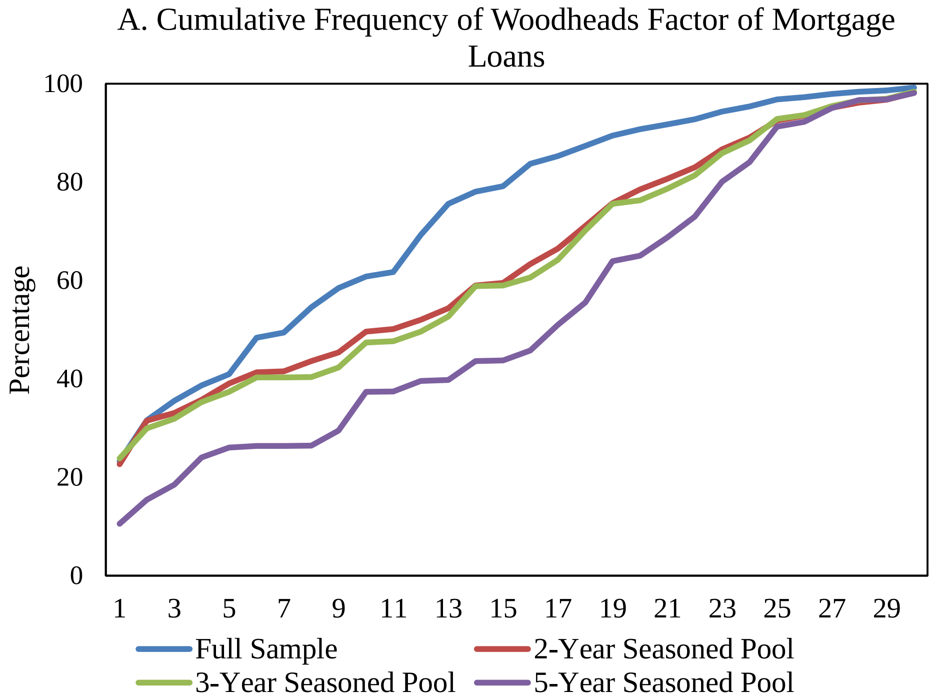
Table 3–3 Descriptive Statistics of Woodheads Factor

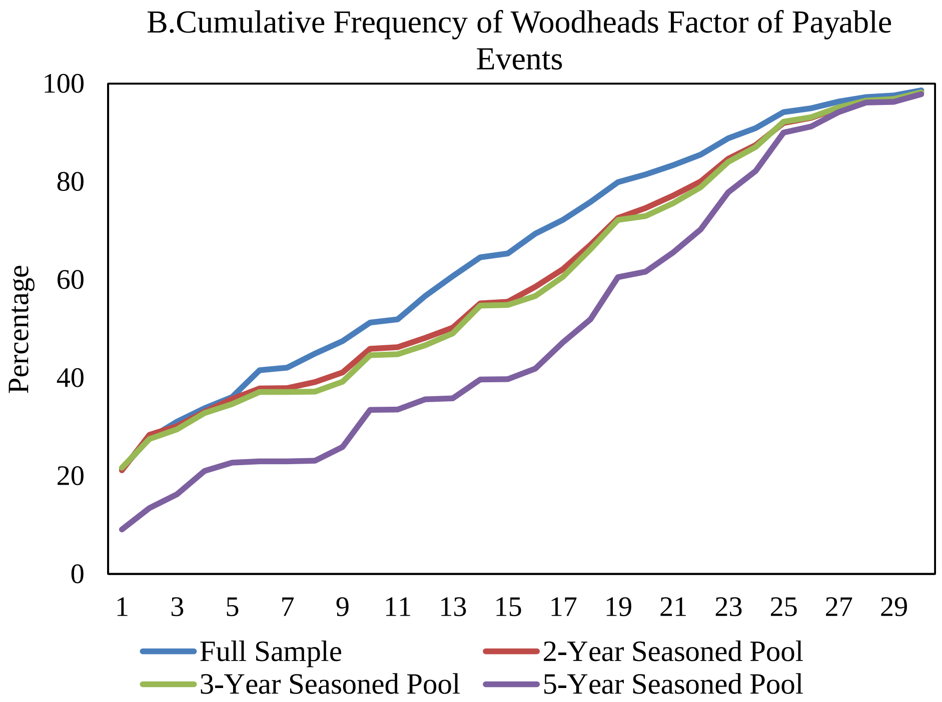
	Full sample	2-Year Seasoned Pool	3-Year Seasoned Pool	5-Year Seasoned Pool
PANEL A- MORTGAGE LOANS				
M=0	14,777	10,522	9,958	6,541
	(8.57)	(11.65)	(14.97)	(27.04)
M=1-2	42,465	17,860	11,399	2,724
	(24.64)	(19.78)	(17.14)	(11.26)
M=3-4	9,534	2,375	2,022	1,516
	(5.53)	(2.63)	(3.04)	(6.27)
M=5-8	21,324	4,484	1,940	431
	(12.37)	(4.97)	(2.92)	(1.78)
M=9-12	19,799	4,748	3,518	2,318
	(11.49)	(5.26)	(5.29)	(9.58)
M≥13	64,429	50,306	37,683	10,657
	(37.39)	(55.71)	(56.65)	(44.06)
Total	172,328	90,295	66,520	24,187
PANEL B- PAYABLE EVENTS				
M=0	482,193	463,346	453,492	309,708
	(9.13)	(11.06)	(12.96)	(19.33)
M=1-2	1,076,621	799,369	630,044	173,871
	(20.38)	(19.08)	(18.00)	(10.85)
M=3-4	237,290	131,740	121,917	98,136
	(4.49)	(3.14)	(3.58)	(6.13)
M=5-8	434,224	171,014	99,155	26,213
	(8.22)	(4.08)	(2.83)	(1.64)
M=9-12	454,945	252,649	217,722	162,397
	(8.61)	(6.03)	(6.22)	(10.14)
M≥13	2,596,909	2,371,427	1,977,673	831,609
	(49.16)	(56.60)	(56.50)	(51.91)
Total	5,282,182	4,189,545	3,500,003	1,601,934

Note: Panel A presents the percentage distribution of M among mortgage loan pools. Panel B presents the same percentage statistics of Panel A for mortgage payable events.

Figure 3-5 presents the cumulative frequency of M among mortgages in these different pools separately for mortgage loans and mortgage payable events.

Figure 3-5 Cumulative Frequency of Woodheads Factor





Note: Panel A presents the cumulative frequency distribution of M among mortgage pools, and panel B presents the cumulative frequency distribution of M among payable events.

Table 3-4 lists the descriptive statistics for the learning factors: both 'Self-Learning Factor' and 'Factor of Learn from Others'. Panel A shows both the frequency and percentage for the self-learning factor. The learning times are calculated as the number of times the borrower has made partial prepayments before the current decision time. 92.48% of borrowers do not have any prior learning in the full sample. In contrast, 66.85% of the borrowers who do not have learning experiences are among those who chose partial prepayment. In other words, more of the individuals who have chosen partial prepayment have experience with making partial prepayments compared with the full sample. 5.61% of the borrowers have had one-time learning experience in the total sample. In contrast, for borrowers who choose partial prepayment, 20.39% of them have had one-off learning experiences, which is much higher than the total sample. It is clear that the partial prepayment experiences of borrowers who choose partial prepayments in the earlier stages are more than the total sample. Panel B shows both the frequency and percentage of factor of learning from others. A similar pattern can be found for the self-learning factor.

Table 3–4 Descriptive Statistics of Learning Factors

Panel A. Self-Learning Factor			Panel B. Factor of Learning from others		
Learning Times	Total Sample	Borrowers who choose Partial Prepayment	Learning Times	Total Sample	If borrower choose Partial Prepayment
0	4,884,687 (92.48%)	19,369 (66.85%)	0	5,232,552 (99.06%)	28,322 (97.76%)
1	296,373 (5.61%)	5,906 (20.39%)	1	29,980 (0.57%)	366 (1.26%)
2	68,739 (1.30%)	2,170 (7.49%)	2	10,789 (0.20%)	153 (0.53%)
3	21,055 (0.40%)	821 (2.83%)	3	4,879 (0.09%)	67 (0.23%)
4	6,692 (0.13%)	357 (1.23%)	4	2,136 (0.04%)	36 (0.12%)
5	2,541 (0.05%)	169 (0.58%)	5	901 (0.02%)	12 (0.04%)
6	971 (0.02%)	80 (0.28%)	6	363 (0.01%)	9 (0.03%)
7	502 (0.01%)	45 (0.16%)	7	227 (0.00%)	2 (0.01%)
8	278 (0.01%)	21 (0.07%)	8	98 (0.00%)	1 (0.00%)
9	192 (0.00%)	11 (0.04%)	9	128 (0.00%)	2 (0.00%)
>=10	152 (0.00%)	23 (0.08%)	>=10	122 (0.00%)	2 (0.00%)
Total	5,282,182	28,972	Total	5,282,182	28,972

Note: The learning times are calculated as the total partial prepayment frequencies of the borrower has been made in the earlier stage before current decision time.

3.4.5 Results

A conditional fixed effects multinomial logit model (FEMNL) is employed to study the risk of mortgage partial prepayments and the process during which mortgage borrowers learn to make partial prepayment decisions in the residential mortgage market in China. Model 1 to Model 3 in Table 3-5 report the regression coefficients and odds ratios in the full sample analysis for partial prepayment choice.

Model 1 is the basic regression with option variables, loan information and borrower characteristics variables. The results show that financial motivation is still important in generating a borrower's partial prepayment decision. The 'call option' is positive and significant, which shows that the alternative investment opportunities, such as stock market investments, are important in explaining a borrower's partial prepayment behavior. For borrowers with two houses and one mortgage, the probability of partial prepayments is higher, the reason being that the pressure they face to make mortgage payments is lower than that faced by borrowers with one house and one mortgage. Borrowers with a graduate degree and above make partial prepayments more often. Monthly income has a negative effect on partial prepayment behavior, possibly because those with higher incomes have greater investment opportunities and they will choose the most profitable one. In China, according to the regulations of lending banks, the minimum amount of partial prepayment is 10,000 Chinese Yuan each time. Borrowers with other income greater than 10,000 Chinese Yuan are less likely to make partial prepayments. The explanation is similar to that of monthly income. The probability of old people to make partial prepayment is lower than young people. Moreover, borrowers with a mortgage on the first house are less likely to make partial prepayments during the payment term. Male borrowers are less likely to partially prepay. In addition, it is easier for mortgage loans with higher loan amounts to be partially prepaid by borrowers. In contrast to the loan quantum, the relationship between loan to value ratio and the probability of partial prepayments is negative. Comparing the results for loan quantum and loan to value ratio, it can be inferred that large loan amounts are accompanied by large housing values instead of higher loan to value ratios.

Model 2 extends Model 1 by adding a woodheads factor into the model, similar to Deng and Quigley's (2012) prepayment model. The woodheads factor M is very significant in accounting for unobserved heterogeneity in this way, increases the magnitude of the option-related variables, and improves the model fit. The negative relationship between M and the probability of partial prepayments indicates that with more missed partial prepayment opportunities (larger M), the probability of making partial prepayments is lower. This is consistent with the burnout effect, which states that sensitive borrowers make partial prepayments as soon as possible, and only the least sensitive borrowers remain in the pool while partial prepayment rates decay.

Model 3 extends Model 2 by adding the 'self-learning factor' into the model. The 'self-learning factor' is positive and significant after controlling for other loan and borrower characteristics and the woodheads factor. This 'self-learning factor' is indicative of a borrower's earlier partial prepayment experiences. The positive relationship between 'self-learning factor' and the possibility of partial prepayment indicates that path dependency exists, since a borrower's partial prepayment decision depends not only upon current stage variables (like other investment opportunities), but also the learning experience of the path. The probability that borrowers with more partial prepayment experiences at earlier stages will make the same decision in the future increases by 26.9 percentage points.

Table 3–5 Self-Learning and Mortgage Partial Prepayment Decisions

Variables	Model1	Model2	Model 3
Woodheads Factor		-0.186***	-0.156***
		(0.010)	(0.010)
		[0.831]	[0.856]
Self-Learning Factor			0.228***
			(0.002)
			[1.269]
Call option	0.109***	0.137***	0.151***
	(0.007)	(0.007)	(0.007)
	[1.115]	[1.146]	[1.163]
Put option	-0.211***	-0.170***	-0.120***
	(0.007)	(0.006)	(0.007)
	[0.809]	[0.844]	[0.887]
2 houses and 1 mortgage	0.025***	0.031***	-0.018**
	(0.008)	(0.008)	(0.008)
	[1.025]	[1.031]	[1.018]
2 houses and 2 mortgages	-0.044***	-0.044***	-0.041***
	(0.007)	(0.007)	(0.007)
	[0.957]	[0.957]	[0.960]
Graduate and above	0.149***	0.157***	0.130***
	(0.006)	(0.006)	(0.006)
	[1.161]	[1.163]	[1.139]
Log monthly income	-0.165***	-0.163***	-0.149***
	(0.006)	(0.006)	(0.006)
	[0.848]	[0.849]	[0.862]
With Other Income more than 10000	-0.360***	-0.352***	-0.250***
	(0.056)	(0.056)	(0.057)
	[0.698]	[0.703]	[0.778]
Age>=40	-0.061***	-0.062***	-0.062***
	(0.006)	(0.006)	(0.006)
	[0.941]	[0.940]	[0.940]
First house	-0.013**	-0.007	-0.003
	(0.006)	(0.006)	(0.006)
	[0.987]	[0.993]	[0.997]
Married	0.046***	0.046***	0.038***
	(0.006)	(0.006)	(0.007)
	[1.047]	[1.047]	[1.038]
With Zhi Cheng	0.049***	0.047***	0.033***
	(0.006)	(0.006)	(0.006)
	[1.051]	[1.048]	[1.034]
Male	-0.011*	-0.011*	-0.009
	(0.006)	(0.006)	(0.006)
	[0.989]	[0.989]	[0.991]
Dependant	-0.008	-0.005	-0.064
	(0.026)	(0.016)	(0.062)
	[0.992]	[0.995]	[0.938]
LTV	-0.027***	-0.042***	-0.058***
	(0.007)	(0.007)	(0.007)
	[0.974]	[0.958]	[0.944]
Log loan amount	0.355***	0.357***	0.312***
	(0.008)	(0.008)	(0.008)
	[1.426]	[1.429]	[1.366]

Year Fixed Effects	Y	Y	Y
Spatial Fixed Effects	Y	Y	Y
Observations	5,207,421	5,207,421	5,207,421
AIC	723012.83	723372.59	703150.00
SC	725315.45	725675.21	705816.19
-2 Log Likelihood	722670.83	723030.59	702754.00

Note: The data is panel dataset, with one observation for each month for each loan during the observation period. The final sample includes 5,207,421 monthly payment events for 172, 328 loans. Multinomial logistic regression is estimated for default, full prepayment and partial prepayment. The results reported here are for partial prepayment. The definitions of the independent variables are shown in Table 2-1. City and origination year fixed effects are included in the regression but not reported. Standard errors are reported in parentheses, and the estimated odds ratios for Multinomial Logit regression are reported in brackets.
*Significant at the 10% level;
** Significant at the 5% level;
*** Significant at the 1% level.

Table 3-6 shows the regression results of learning from others and prior mortgage partial prepayment decisions. Compared to Table 3-5, one factor named “Learning from Others” is added into the model. The definition of this variable can be found in the previous section. The coefficients for both the self-learning and the learning from others factor are positive and significant, after controlling for other loan and borrower characteristics and the woodheads factor. For borrowers who learn from their own experiences in earlier stages, the probability that they will make the same decision in the future increases by 26.2 percentage points. The probability that borrowers who learn from the experiences of others in earlier stages will make the same decision in the future increased 1.8 percentage points. For the factor of learning from others, this positive relationship can be explained as such: borrowers can learn from their colleagues or friends from the same company. The more colleagues or friends who have partially prepaid in the past, the higher the probability that borrowers choose to partial prepay on their own mortgage.

Table 3–6 Learn from Others and Mortgage Partial Prepayment Decisions

Variables	Model 4
Woodheads Factor	-0.170***
	(0.010)
	[0.844]
Self-Learning Factor	0.233***
	(0.002)
	[1.262]
Learning From Others	0.017***
	(0.003)
	[1.018]
Call option	0.107***
	(0.007)
	[1.113]
Put option	-0.076***
	(0.006)
	[0.927]
2 houses and 1 mortgage	0.022***
	(0.008)
	[1.022]
2 houses and 2 mortgages	-0.041***
	(0.007)
	[0.960]
Graduate and above	0.130***
	(0.006)
	[1.138]
Log monthly income	-0.147***
	(0.006)
	[0.863]
With Other Income more than 10000	-0.247***
	(0.057)
	[0.781]
Age>=40	-0.055***
	(0.006)
	[0.947]
First house	-0.006
	(0.006)
	[0.994]
Married	0.040***
	(0.007)
	[1.041]
With Zhi Cheng	0.034***
	(0.006)
	[1.034]
Male	-0.008
	(0.006)
	[0.992]

Dependant	-0.052
	(0.062)
	[0.950]
LTV	-0.087***
	(0.007)
	[0.917]
Log loan amount	0.307***
	(0.008)
	[1.360]
Year Fixed Effects	Y
Spatial Fixed Effects	Y
Observations	5,207,421
AIC	714370.42
SC	716753.83
-2 Log Likelihood	714016.42

Note: The data is panel dataset, with one observation for each month for each loan during the observation period. The final sample includes 5,207,421 monthly payment events for 172, 328 loans. Multinomial logistic regression is estimated for default, full prepayment and partial prepayment. The results reported here are for partial prepayment. The definitions of the independent variables are shown in Table 3-1. City and origination year fixed effects are included in the regression but not reported. Standard errors are reported in parentheses, and the estimated odds ratios for Multinomial Logit regression are reported in brackets.
*Significant at the 10% level;
** Significant at the 5% level;
*** Significant at the 1% level.

Recent experience may play a larger role than old experience in determining behavior. Here, the results of the recency effect for self-learning are listed next. Table 3-7 shows the regression for recency effect and mortgage partial prepayment behavior. Only those who have been partial prepaid on their own mortgages at least once are selected. A factor termed ‘Self-Learning Duration’ is added into the model. It calculated as the duration between current time and the time of latest partial prepayment decision.If the ‘Self-Learning Duration’ is large, then the partial prepayment experience is older, and vice versa. If the recency effect exists, we would expect a negative relationship between ‘Self-Learning Duration’ and mortgage partial prepayment probability. In other words, partial prepayment probability would be smaller as the time from the previous partial prepayment decision to current time is longer. From Table 3-7, it can be seen that the 'Self-Learning Duration' is negatively and significantly related to partial prepayment probability as expected.

Table 3–7 Recency Effect and Mortgage Partial Prepayment Decisions

Variables	Model 5
Woodheads Factor	-0.152***
	(0.011)
	[0.863]
Self-Learning Factor	0.029***
	(0.007)
	[1.023]
Learning From Others	0.014**
	(0.006)
	[1.014]
Self-Learning Learning Duration	-0.676***
	(0.013)
	[0.509]
Call option	0.078***
	(0.007)
	[1.080]
Put option	-0.102***
	(0.007)
	[0.903]
2 houses and 1 mortgage	0.026***
	(0.008)
	[1.025]
2 houses and 2 mortgages	0.003
	(0.006)
	[1.003]
Graduate and above	0.018***
	(0.006)
	[1.018]
Log monthly income	-0.024***
	(0.007)
	[0.975]
With Other Income more than 10000	-0.000
	(0.006)
	[1.000]
Age>=40	0.008
	(0.006)
	[1.008]
First house	-0.002
	(0.007)
	[0.998]
Married	0.012*
	(0.007)
	[1.013]

With Zhi Cheng	0.013**
	(0.006)
	[1.013]
Male	-0.005
	(0.006)
	[0.995]
Dependant	-0.020***
	(0.007)
	[1.020]
LTV	0.007
	(0.008)
	[1.007]
Log loan amount	0.123***
	(0.009)
	[1.131]
Year Fixed Effects	Y
Spatial Fixed Effects	Y
Observations	759,152
AIC	292258.63
SC	294335.82
-2 Log Likelihood	291898.63

Note: Only loans which have been partially prepaid are included. The data is panel dataset, with one observation for each month for each loan during the observation period. The final sample includes 759,152 month payment events. Multinomial logistic regression is estimated for default, full prepayment and partial prepayment. The results reported here are for partial prepayment. The definitions of the independent variables are shown in Table 3-1. City and origination year fixed effects are included in the regression but not reported. Standard errors are reported in parentheses, and the estimated odds ratios for Multinomial Logit regression are reported in brackets.
*Significant at the 10% level;
** Significant at the 5% level;
*** Significant at the 1% level.

3.4.6 Alternative Argument

The previous sections have documented that the probability of partial prepayments is explained by the learning process. However, there is an alternative argument: mortgage borrowers in China choose to make partial prepayments on their mortgage loans when they have earnings surprises (extra income). This would mean that past experience in making partial prepayments may simply be a proxy of the likelihood that the borrower has received extra income. Being in the same company as someone else who has partial prepayment experience can be another proxy of that likelihood – borrowers working in the same company have a good chance of receiving bonuses at the same time. Therefore, it is important to distinguish such 'earning surprises' from the learning experience.

The results of the recency effect shown in Table 3-7 can help to respond to this alternative argument. Following the extra income argument, if the 'Self-Learning Duration' is large, then the probability of getting extra income is high, and the probability of mortgage partial prepayments is high (positive effect). However, the results are in the opposite direction.

3.6 Conclusion

This chapter studies the risk of partial prepayment and the reinforcement learning process of borrowers' partial prepayment behavior in the Chinese mortgage market. The results indicate that the risk of partial prepayment is different from the full prepayment risk. Borrowers' partial prepayment behavior follows the reinforcement learning process.

The residential mortgage market in China is an important financial engine for the booming housing market. As the Chinese mortgage market is very different from that of the US, especially in terms of the motivation to make prepayments and the calculation of call options, the option theory is not applicable in predicting the risk of partial prepayments in China. Since the stock market provides a higher return on investment in the capital market for

Chinese households, fluctuations in the stock market have a more significant impact on the probability that borrowers will terminate their mortgages, especially by making prepayments.

The characteristics of borrowers have a significant impact on their propensity to make partial prepayments, and thus may useful for screening loan applicants and determining potential high-risk borrowers. The results in this chapter also indicate that the partial prepayment behavior of borrowers is path-dependent and follows a reinforcement learning process. The partial prepayment decision depends not only upon current stage variables and borrowers' characteristics, but also learning experiences, both from their own experiences and from others). Borrowers who have made more partial prepayments early on are more likely to continue making partial prepayments compared to those who have less experience with making them. In addition, recency effects are also found in the self-learning process.

Chapter 4 Predicting Default of Chinese Companies: Information beyond Accounting and Market Variables

4.1 Introduction

Predicting the default risk of corporations is a widely-studied issue that is critical for credit risk management, macroeconomic policy-making and financial regulation. Assessing which variables are relevant in predicting the default risk of firms is an important issue, not only to providers of capital, but to academics and economists as well. However, it is a challenging task. This is especially true for developing countries where the quality of information is poor. This chapter builds default probability prediction models for Chinese companies that address this information quality issue.

In studies in the United States, discriminant analysis using accounting ratios has long proven to be a useful tool for predicting corporate bankruptcy (Altman 1968; Ohlson 1980; Zmijewski 1984; Beaver, McNichols, and Rhie 2005; Altman 2012). The accounting data is derived from the firm's operating environment, and includes variables such as firm size, book-to-market equity ratio, and its rate of growth. The rationale for this approach is that historical accounting information is a good predictor of firms' future insolvency.

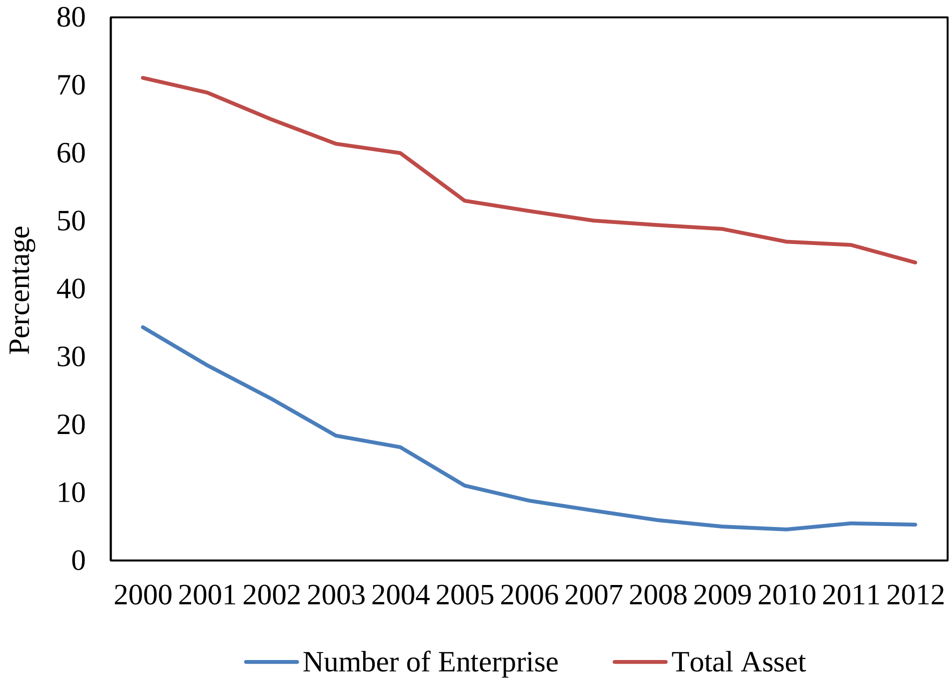
Recently, researchers have suggested that market-driven variables derived from a firm's information and trading environments be used instead of accounting ratios to predict corporate bankruptcy. They argue that the probability of a firm entering default should be perfectly reflected at all times in the market value of its equity (Chava and Jarrow 2004; Hillegeist et al. 2004). If efficient market prices reflect the full forward-looking information of a firm's performance and thus its likelihood to enter bankruptcy, accounting-based variables should not contain any information that is not reflected in the market price of a company. Several studies have tried to combine these two approaches to predict corporate default (Shumway 2001; Campbell, Jens, and Jan 2008).

However, in the context of developing countries, information quality becomes a potential issue when either the accounting-based model or the market-driven model is used to predict corporate default. First, accounting and legal standards in developing countries are often less stringent and detailed than those in developed countries. In their study on the different level of protection investors received from capital markets in various countries, La-Porta et al. (1997) found that poor accounting and legal standards were a reflection of the size and maturity of a country's capital markets. Developing countries, including some in Asia, provide investors with less protection, and have weaker capital markets than those with good legal environments. Secondly, as mentioned in Shleifer (1994): "Politicians can shut down a business, kick it out of its premises, or even refuse to allow it to start". This institutional effect in developing counties cannot be overlooked. Morck, Yeung, and Yu (2000) showed that institutional effects played a unique role in several Asian markets, leading stock prices to be less informative in these markets.

As a leading emerging economy, China has witnessed all of the concerns raised above. Thus, using accounting and market variables to predict the default probability of Chinese-listed companies is an urgent question. China has a unique corporate environment where a large proportion of state-owned enterprises (SOEs) exist and have significant roles in the market. SOEs are usually alleged to make distorted production and operation decisions because of a mixture of poor governance, under-developed management skills, and heavy "policy burdens" to fulfil government policy agendas (Lin and Tan 1999; Lin and Li 2008). However, SOEs and other companies with a high level of state ownership also have privileged access to inputs, capital, and markets which have a positive effect on their performances. More importantly, these companies usually have extended social capital, such as bank lending relationships, government support, and subsidies. What is the impact of these factors on corporate default? This chapter is especially interested in the impact of state ownership on company default after controlling for variations in observable firm financial performances measured by firm-level accounting and market variables.

Figure 4-1 shows the share of SOEs in different industrial sectors, both by numbers of enterprise and total assets. There is a substantial gap between these two lines, indicating that even though there are only a few SOEs, their total asset are much larger than non-SOEs. Again, the data suggests that the average size of SOEs is much bigger than non-SOEs. In 2000, even though SOEs were less than 40% of the total number of enterprises, they accounted for more than 70% of total assets. Moreover, SOEs' share of industry has declined enormously; their shrinking share in the industrial sector both in terms of the number of enterprises and total assets is also evidenced in Figure 4-1. This is due to the stronger growth of non-SOEs as the economy has developed.

Figure 4-1 Share of SOEs in the Industry Sector



Note: From National Bureau of Statistics. The Y-axis measures the percentage, while the X-axis indicates the year.

The significance of macroeconomic variables in default prediction is the third issue which being investigated in this chapter. Nichell, Perraudin, and Varotto (2000) and Bangia et al. (2002) showed that changes in ratings in the US were affected by the state of the economy. However, the usefulness of macroeconomic variables for assessing default is not limited to their impact on firm-specific accounting and market variables. For example, the interdependence of firms and the potential contagion of default risk may cause the default risk of individual firms to exceed the level that is warranted or can be controlled by a particular firm. Also, the aggregation of data may neutralize the noise contained in firm-level accounting and market variables. This essay will test whether macroeconomic variables provide additional information beyond firm-specific accounting and market information in assessing the probability of corporate default.

Finally, this chapter explores how the default risk of Chinese companies is correlated and clustered. Lennox (1999), Chava and Jarrow (2004), and Moody's (2006) showed significant industry fixed-effects in the bankruptcy of US firms after controlling for accounting and market variables, suggesting that bankruptcy risk was clustered by industry and that intra-industry bankruptcy risk was correlated because of common observable risk factors. Similar industry fixed-effects will be tested in this chapter. It is well understood that when information quality is poor, risk factors are likely to become latent. Therefore, in an alternative specification, instead of fixed-effects, industry random effects will be tested. That implies that while default risk is clustered by industry, the correlation of intra-industry default risk is due to common unobservable risk factors. The test is conducted by estimating a shared frailty model where cluster effects are incorporated into the model as independent and identically distributed random variables.

The main data of this chapter is from National University of Singapore (NUS) Risk Management Institute's (RMI) corporate default database. After data cleaning, there are 1,897 public listed non-financial companies in the sample. During the period from 1994 to 2010, 201 firms had defaulted. Through this database, quarterly financial statements and daily stock price information for all these firms can be obtained, as well as interest rate and macroeconomic variables during the study period. This chapter also elaborates the RMI database by additional variables such as firm-specific ownership structure and China's effective exchange rate.

Firstly, the default probability of Chinese companies will be assessed using a standard Cox proportional hazards model with only firm-level accounting and market variables. The results indicate that several variables, such as distance-to-default (DTD), cash to total asset ratio and net income to total asset ratio, have a significant impact on the possibility of corporate default, despite the concern that accounting variables can be easily manipulated in China and that market variables are noisy because of market inefficiency. In the next version of the model, state ownership and macroeconomic variables, such as short-term bank lending rate, effective exchange rate, growth in money supply, the coincident index and inflation, are added. Interestingly, the results show that state ownership has a strong negative impact on corporate default, confirming the conjecture that state ownership is overall a benefit to Chinese companies. The macroeconomic variables also strongly predict corporate default. After adding these variables, the accounting and market variables remain significant and the model fit is improved. The cumulative accuracy profiles (CAP curves) of the two sets of models are compared and the results indicate that including state ownership and macroeconomic variables increases within-sample predictive power.

Following the Bloomberg industry classification, all corporations in the sample are classified into nine categories. The raw default rates are significantly different among different industries, with the consumer-cyclical industry having the highest default rate while the energy and utilities industry has the lowest default rate. In a fixed-effects model, the risk of corporations entering default is significantly higher in certain industries, such as communications, diversified and technology, after controlling for accounting, market, ownership and macroeconomic variables. In the shared frailty model, statistically significant random effects are allowed, suggesting that the risk of default is correlated within each industry because of common unobservable risk factors.

The model developed here is useful for risk assessment and management. This chapter also has several implications for investors. First, financial statements and stock price and volatility provide useful but insufficient information about the health of Chinese companies. Thus, lenders and investors should be aware of other factors that affect their lending and investment risk. From a portfolio management perspective, macroeconomic factors are more difficult to diversify away. Second, SOEs differ from non-SOEs not only in observable ways (through their financial statements and stock market indicators), but also in non-observable ways (as reflected by their lower marginal default risk). The exact mechanism through which state ownership helps SOEs avoid default deserves further research. Finally, the default risk of Chinese companies is clearly clustered, as suggested by the industry fixed-effects or the industry random effects that were found. However, if the intra-industry default correlation is indeed due to common latent factors as the random effects model suggests, then that source of default risk cannot be diversified away. The time-series characteristics of the frailty factor should be further studied, for example by following Duffie et al. (2009), to better understand portfolio default risk.

This chapter has clear policy implications. For example, it found that the effective exchange rate is negative related to corporate default, which suggests that depreciating the Chinese RMB would not only reduce the profitability of many Chinese companies but also significantly increase the default risk of corporations. Money supply, bank lending rate and inflation also have significant impacts on corporate default, suggesting that policy makers should take into consideration the impact of monetary policy on corporate default.

The rest of this chapter proceeds as follows: in section 4.2, the relationship between this chapter and prior literature are presented; section 4.3 show the data and empirical methodology; the comprehensive empirical results are presented in section 4.4; and concluding remarks are in the final section.

4.2 Related Literature

The credit risk of corporations is very well-studied and several corporate bankruptcy prediction models have been developed based on different methods. Some well-known studies attempt to predict corporate default using accounting-based variables. After Beaver (1966), many studies predicting corporate default based on accounting ratios quickly appeared. Beaver (1966) applied a business failure prediction model based on financial ratios and showed that certain financial ratios gave statistically significant signals for judging corporate failure. Altman (1968) argued that the Univariate Discriminant Analysis used by Beaver (1966) could be confusing in predicting failure, and extended Beaver (1966)'s analysis by combining several measures in a Multivariate Discriminant Analysis (MDA) model. Altman et al. (1977) improved their bankruptcy classification model to a more accurate Zeta analysis that could predict corporate insolvency up to five years before the event.

Since the 1980s, the Multivariate Discriminant Analysis (MDA) model has been criticized on the grounds that its requirements are restrictive. Some other methods, such as logit Analysis (LA) and probit Analysis (PA), have emerged to overcome the limitation of the MDA model. Ohlson (1980) was the first to apply logit regression method to predict corporate failure. He used conditional logit models to predict the probability of failure for US firms, and was followed by Menash (1984). Helfert (1982) developed the cash components model to predict corporate default. Genry, Newbold, and Whitford (1985) redesigned the Helfert (1982) cash components model and adopted the logit regression method for their analysis. Besides the logit regression model, other studies have used probit analysis for analysing failure. Zmijewski (1984) criticized the sampling approach of the logit regression method, and developed a weighted probit bankruptcy prediction model.

All of these approaches predict future bankruptcy based on accounting ratios drawn from a firm’s financial statements. While accounting ratios are important for predicting corporate default, and models based on accounting ratios provide a good framework for predicting corporate default, there are several criticisms regarding the use of accounting ratios for predicting failure. First, accounting ratios tend to look backward and reflect past performance. Thus, they are not effective for predicting the future status of companies, which means their ability to predict corporate future bankruptcy is questionable. Second, accounting measures underestimate the book value of assets because of accounting conventions (e.g. historical cost) (Charalambakis, Espenlauby, and Garrett 2009). Third, sometimes, accounting cannot reflect the real performance of corporations, because accounting data can be manipulated by management (Agarwal and Taffler 2008a).

To overcome the shortcomings of accounting ratios, researchers began to include market-driven variables in their models. Compared with backward-looking accounting ratios, the information in stock prices tends to be more forward looking. Information that may not be contained in accounting statements could be reflected in the price of stocks if the companies are listed and frequently traded. Black and Scholes (1973), and Merton (1974) established a market-based approach by using an option pricing method. A firm's capital structure is assumed to be composed by equity and a zero-coupon bond with maturity T and face value of D. The firm's equity is simply a European call option with maturity T and strike price D on the asset value and, therefore, the firm's debt value is just the asset value minus the equity value. Under this model, default can occur only when the debt matures, that is, when the firm's assets are no longer sufficient to cover debt obligations, a scenario that is odds with reality. Black and Cox (1976) relaxed this assumption and proposed that default may occur any time between the issuance and maturity of the debt when the firm's assets are no longer sufficient to cover debt obligations. Several studies, such as Crosbie and Bohn (2002), Hillegeist et al. (2004), Vassalou and Xing (2004), Bharath and Shumway (2008), Reisz and Perlich (2007), further developed this market-based concept to determine the possibility of corporate failure. Crosbie and Bohn (2002) extended the contingent claims framework approach and transformed distance-to-default into an expected default frequency (EDF) using an empirical default distribution. Vassalou and Xing (2004) computed the default likelihood indicator (DLI) to measure bankruptcy probability in such a framework. Hillegeist et al. (2004) used a similar approach to compute bankruptcy probability and found that the BSM-probability methodology outperforms both Altman’s Z-score and Ohlson’s O-score in bankruptcy prediction. The difference between Vassalou and Xing (2004) and Hillegeist et al. (2004) is a technical issue around adjustments for dividends.

There are a number of problems for using market-based models to predict bankruptcy. As Saunders and Allen (2002) argue, these models are based on some unusual assumptions, such as the non-normality of stock returns and the similarity of all debts. Some papers have attempted to combine accounting and market data to predict corporate default (Shumway 2001; Campbell et al. 2008).

Both accounting-based and market- driven variables play an important role in predicting corporate default, and substantial research has been done to improve the models that use kinds of variables. However, when these models are applied to developing countries, especially China, information quality becomes a potential problem, which should be a key concern. There is an urgent need to explore information beyond accounting and market variables to predict the default of Chinese companies.

4.3 Data and Empirical Methodology

4.3.1 Data

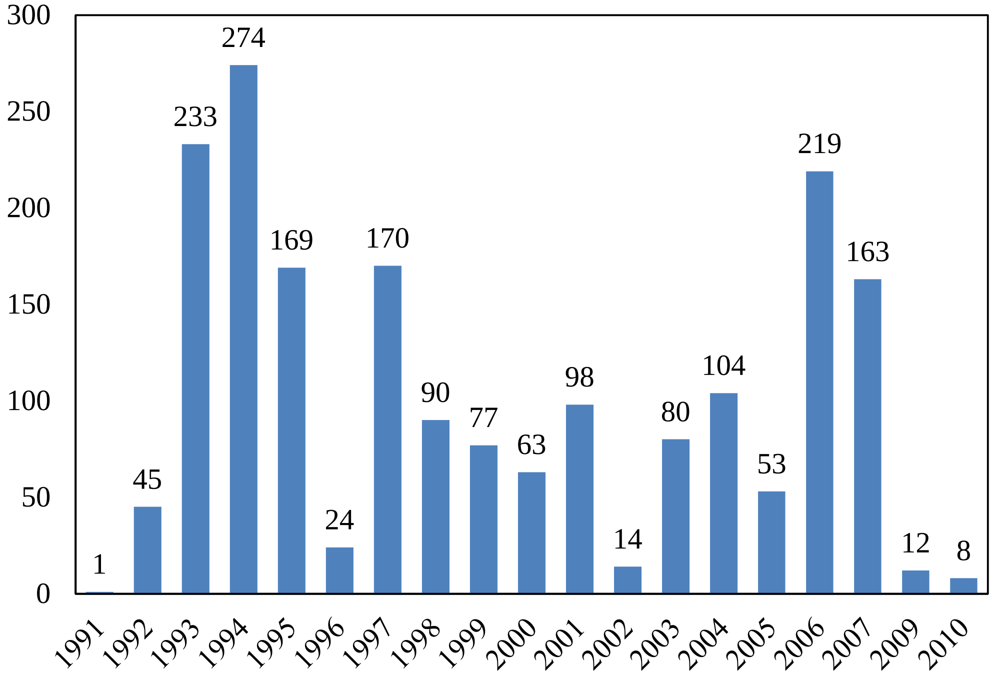
The data used in this chapter is from the NUS RMI's corporate default database. [23] This database contains the historical observations of all corporations that have been listed on the two Chinese stock exchanges, the Shanghai Stock Exchange and the Shenzhen Stock Exchange, from 1991 to 2010. Over this period, 2,063 companies were listed during this period[24]. After excluding financial companies, 1,897 companies were left in the sample.[25] Figure 4-2 depicts the number of new listings in each year, that is, the number of new companies that entered the sample each year. There were a large number of new listings between 1993 and 1995 and between 2006 and 2007. 274 companies were listed in 1994 alone. The number of new listings declined substantially after the Asian Financial Crisis in 1998, as well as after the recent Great Recession.

[23] More information on this database is available at: http://www.rmi.nus.edu.sg/

[24] About 3 percent of the companies were excluded from the analysis because of missing data.

[25] To implement Merton's model and calculate Distance to Default (DTD), the so-called KMV assumption is typically adopted. This assumption sets T-t to one year and L to the firm's book measure of short-term debt plus one half of its long-term debt. The KMV implementation assumption becomes problematic for financial firms. Financial firms typically have a large amount of liabilities that are neither classified as short-term nor long-term debt, and thus the KMV assumption would grossly understate the amount of debt. Therefore, financial firms are excluded from the sample.

Figure 4-2 Number of New Listings by Year



Note: Newly listed companies are those enter the sample each year. The Y-axis measures the company number, while the X-axis indicates the year.

The NUS RMI database includes information on corporate events for all the companies in the sample. Corporate events include bankruptcies, default, debt offerings, mergers and acquisitions, stock repurchases, stock splits, and changes of industry. Default is defined as one of the following: bankruptcy, coupon and/or principal payment default, loan payment delay, loan covenant violation, debt restructuring, and subsidiary default. Out of the 1,627 corporate events in the sample, only a small fraction is default.

The quarterly financial statements for each company, which contain the necessary accounting information, are obtained. The event histories of all the companies give rise to 77,481 company-quarter observations. The stock price, outstanding shares and market capitalization of each company on each trading day over the study period are also obtained. The indices of the Shanghai Stock Exchange and the Shenzhen Stock Exchange are also included in the data.

Ownership data is collected from RESSET, a data company in China. RESSET traces the ownership structure of each public company over time. We are particularly interested in state ownership, which is calculated as the number of state shares divided by the total number of shares outstanding. It is a time-varying variable, as state shares can be transferred to domestic institutions upon the approval of China's Securities Regulatory Commission although they cannot be traded.

Most of the macroeconomic variables are obtained from the RMI's database. This dataset is augmented by the real effective exchange rate time series from the Bank for International Settlements (BIS).

4.3.2 Empirical Methodology

The default probability model estimated is a standard Cox proportional hazard model that was used by Shumway (2001) to predict corporate bankruptcy, and is used widely in the mortgage default literature (e.g. Quigley and Van Order 1991; Vandell et al. 1993; An. Deng, and Gabriel 2011). The hazard model is convenient mainly because it allows us to work with the full sample of companies despite some observations being censored when collecting the data.

Assume that the hazard rate of default of a company at period T since it enters our sample follows the form:

$$h_i(T;\ X_{it}) = h_0(T)exp(X_{it}'\beta), i = 1, \dots, N, t = 1, \dots, S. \tag{4.1}$$

Here $h_0(T)$ is the baseline hazard function, which only depends on the age (duration), T, of the firm and is an arbitrary function that allows for a flexible default pattern over time[26]; X_{it} is a vector of covariates for each individual firm i that includes all the identifiable default risk factors that can be either time-constant or time-varying (t dependent). In this proportional hazard model, changes in covariates shift the hazard rate proportionally without otherwise affecting the duration pattern of default. The hazard model is estimated with Maximum Likelihood Estimation (MLE) using the firms' event-history data[27].

The first set of covariates included in the model consists of the established accounting and market variables from the existing literature. These include: distant-to-default (DTD), cash to total asset ratio (Cash), net income to total asset ratio (Net Income), market-to-book asset ratio (M/B), idiosyncratic volatility (Sigma), asset growth rate (Mu), and market cap (Size)[28]. Other accounting variables, such as working capital, retained earnings, EBIT, sales, current ratio, and average leverage ratio, are also considered. Since accounting statements are usually released several months after the reporting period, all the accounting variables are lagged by one quarter. This is similar to the treatment in Duan, Sun, and Wang (2010).

The focus in the chapter is on information beyond accounting and market variables. The first such variable considered is ownership structure. Many Chinese companies are either completely state-owned (SOEs) or owned to

---

[26] Notice that the company's live duration time T is different from the natural time t, which allows identification of the model.

[27] Refer to Kalbfleisch and Prentice (1980) for details about the MLE estimation of the hazard model.

[28] Duan, Sun, and Wang (2010) provide detailed definitions of these variables.

some degree by different parts of the state. SOEs are products of China's "socialist transformation" completed in 1956 and were originally government agencies set up in various industrial sectors to fulfil China's "planned economy". After several waves of reform during China's implementation of a "market-oriented economy", most SOEs became shareholder-owned and many are now listed on domestic or international stock exchanges. However, state governments including the central, provincial, municipal and local district governments still have large ownership shares in many of them. The focus in this chapter is on assessing the impact of state ownership after controlling for accounting and market variables. The marginal effect of state ownership possibly exists because SOEs and companies with substantial state ownership usually have extended social capital, such as bank lending relationships, government support, and subsidies.

The macroeconomic variables included in the model are: the short-term bank lending rate, growth in money supply, real effective exchange rate, stock market return, inflation and a coincident index. Firms rely heavily on short-term debt so the short-term bank lending rate is a key variable for their operational costs. Money supply growth is an indicator of the corporate financing environment. China has a huge trade surplus and many companies are export related. Therefore, the exchange rate has a significant impact on Chinese companies. The real effective exchange rate takes into consideration the exchange rate between the Chinese RMB and other major currencies including US dollar and Euro. The return on the stock market is an overall measure of corporate health and profitability, while the coincident index and inflation measure the health of the macro economy. Other macroeconomic variables, including per capita disposable income, real GDP growth, unemployment rate, yield slope, consumer confidence, producer price index, and government spending, are also considered.

The last set of variables considered is industry dummies. Following the Bloomberg classification of industries, all the companies are classified into the following nine industries: Basic Materials, Communications, Consumer Cyclical, Consumer Noncyclical, Diversified, Energy, Industrial, Technology and Utilities[29]. Dummy variables are used in the model to capture industry-fixed effects. The utilities industry is used as the reference group.

Instead of using industry dummies to form a fixed-effects model, industry random- effects using a shared frailty model are considered. In this model, default is clustered by industry and the correlations between defaults in the same industry (cluster) are modelled with a random component for the hazard function. The hazard rate for the i th individual in the j th cluster is

$$(4.2) \qquad h_i(T;\ Z_{it};j) = h_0(T)exp\left(Z'_{it}\beta + \gamma_j\right), i = 1, \ldots, N, t = 1, \ldots, S, \quad j = 1, \ldots, m.$$

Again $h_0(T)$ is the baseline hazard function. Z_{it} is the vector of covariates, which are all the same covariates in X_{it} except the industry dummies. γ_j is the random effect for industry (cluster) j. The random components $\gamma_1, \gamma_2, \ldots \gamma_m$

[29] The whole financial industry is excluded from the sample. Following Daun et al. (2010), to implement Merton's model to calculate the distance to default, the so-called KMV assumption is typically adopted where the short-term debt and long-term debt are used. However, financial firms typically have large amount of liabilities that neither classified as short-term nor long-term debt, and the KMV assumption becomes problematic which grossly understate the amount of debt of financial firms. Based on this consideration, the financial industry is excluded from the regression sample.

are assumed to be independent and identically distributed as a normal random variable with mean 0 and an unknown variance θ. θ will be estimated together with coefficients β with MLE[30].

4.4 Empirical Results

The study period of this chapter is from the first quarter of 1994 (1994Q1) to the last quarter of 2010 (2010Q4)[31]. In Table 4-1, a summary of the total number of companies in the sample as well as the total number of defaults and total number of firms that were delisted during the study period are presented. Among the 1897 non-financial companies, 201 defaulted and 21 were delisted. The cumulative default rate is 10.6% and the cumulative delisting rate is 1.1% during 1994-2010.

Table 4–1 Status Distribution of Sample

Status	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	201	10.6	201	10.60
2	21	1.11	222	11.70
3	1,675	88.30	1,897	100
Total	1,897	100	1,897	100

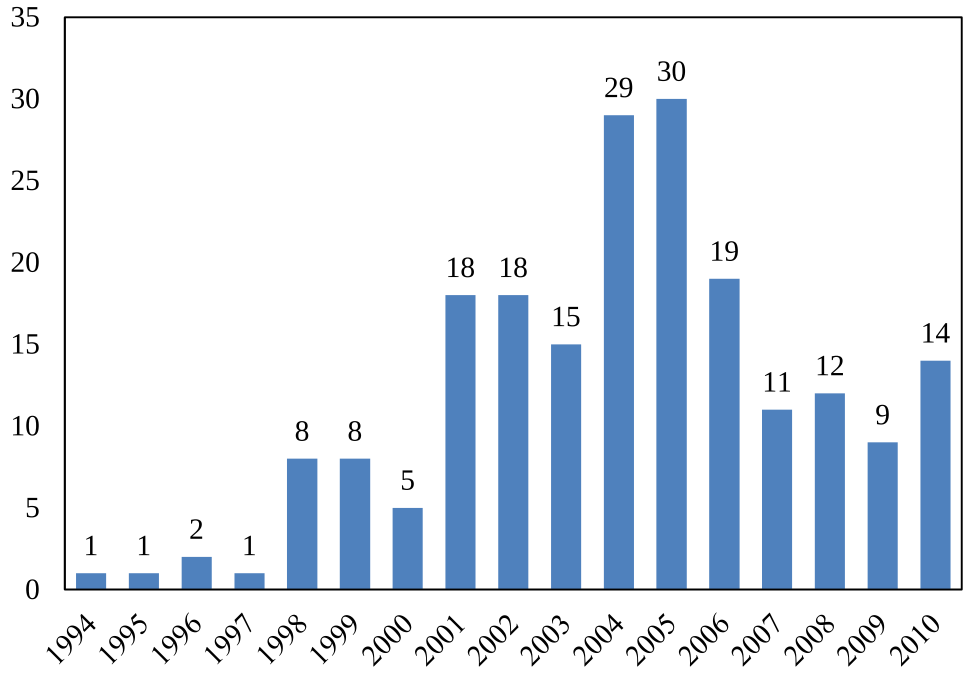
Note: status=1 indicates default; status=2 indicates delisting; and status=3 indicates survival.

The number of defaults in each year is plotted in Figure 4-3. Given the small proportion of delisting in the sample, delisting is not modelled separately. Instead, it is treated as censored.

[30] See Therneau and Grambsch (2000) for more details.

[31] Given the unavailability of many variables, the left censored time in the data is 1994Q1 (some companies were listed before 1994). This creates a left censoring problem. How this left censoring affects the results is assessed and little impact is found.

Figure 4-3 Number of Defaults in Each Year



Note: Number shown here is the default number of companies in each year. There is a clear increase of the number of defaults from 1997, and the peak value appears in 2005, following by 2004. The Y-axis measures the default number, while the X-axis indicates the year.

Preliminary analysis shows that accounting and market variables have unreasonably high variations, possibly due to outliers. All accounting and market variables are winsorized at the 1 and 99 percentiles. In Table 4-2, the descriptive statistics of covariates after winsorization are reported. There are still large variations in the distance-to-default ratio. Most companies in most of the quarters have positive DTD, suggesting they are not near default risk. The average cash-to-total asset ratio is about 15 percent and the average net income-to-total asset ratio is about 4 percent. The average market-to-book ratio (M/B) is almost 4 times, suggesting that some companies are potentially over-valued. Size is the log of market capitalization normalized by real GDP.

Table 4–2 Descriptive Statistics of Variables

Variables	Obs.	Mean	Std Dev	Minimum	Maximum
dtd	65,979	6.2900	5.2449	-0.0428	29.2149
cash	65,925	0.1541	0.1278	0.0036	0.6448
Net income	65,872	0.0417	0.05831	-0.1205	0.2594
m2b	44,457	3.7710	2.9862	0.0591	17.4842
sigma	59,451	1.8025	1.7764	0.0382	10.0085
size	48,499	-2.6215	1.1865	-5.9912	0.3754
State share	77,001	0.2720	0.2900	0	1
Stock return	77,001	0.0054	0.0539	-0.1226	0.1989
CHLDI6M	77,001	6.0004	1.4591	4.86	10.8
m1gr	77,001	0.0137	0.0110	-0.0168	0.0472
REER	77,001	91.6743	7.0523	65.18	105.3667
coincindex	77,001	99.3892	3.2610	93.05	109.4067
inflation	77,001	2.9235	4.8038	-2.1667	26.9

Note: *DTD*: distance-to-default, which is a volatility adjusted leverage measure based on Merton (1974). DTD in this chapter is calculated following Duan et al. (2010), which is described in detail in appendix 1; *Cash*: the sum of cash and short-term investments of corporate; *Net income*: the net income of corporate; m2b: market to book value ratio; *Sigma*: 1-year idiosyncratic volatility, calculated by regressing individual monthly stock return on the value-weighted corresponding stock market monthly return over the preceding 12 months. SIGMA is the standard deviation of the residuals from the regression. Following Shumway (2001), SIGMA is treated as missing if there are less than 12 monthly returns; *Size*: log of the ratio of firm's market equity value to the average market value; *State share*: proxy of state ownership. It is calculated as the ratio of state owned share on total shares; *Stock return*: firm's stock return; *CHLDI6M*: China lending rate of short than 6 month, which is regulated by the People's Bank of China; *m1gr:* proxy of money supply in China; *REER*: real effective exchange rate of China; *Coincindex*: coincident index of China; *Inflation*: inflation rate in China.

In the benchmark default probability model, only firm-specific accounting and market information are included. The estimates of the benchmark hazard model are contained in Table 4-3. All the covariates are significant at the 95% significance level. For example, distance-to-default (DTD) has a negative relation with default – the higher the distance from default, the lower the probability of default. Cash-to-total asset ratio (Cash) and net income-to-total asset ratio (Net Income) are also negatively correlated with default probability. Surprisingly, market-to-book ratio (M/B) has a positive relationship with default while idiosyncratic volatility (Sigma) has a negative relationship with default. In an efficient market, a high M/B implies high growth opportunities and a bright future for a company, and thus lower chance of default. A possible explanation of the surprising result for Chinese companies is that some companies are significantly over-valued and will thus fall into default easily. This is consistent with the claim that some issuers in China use IPO as an opportunity to raise capital that is not used subsequently for production and operations. Finally, size has a significant negative impact on default, meaning that large companies are less likely to default. Other accounting and market variables such as working capital, retained earnings, earnings before interest and taxes (EBIT), sales, and leverage are ruled out due to multicollinearity concerns.

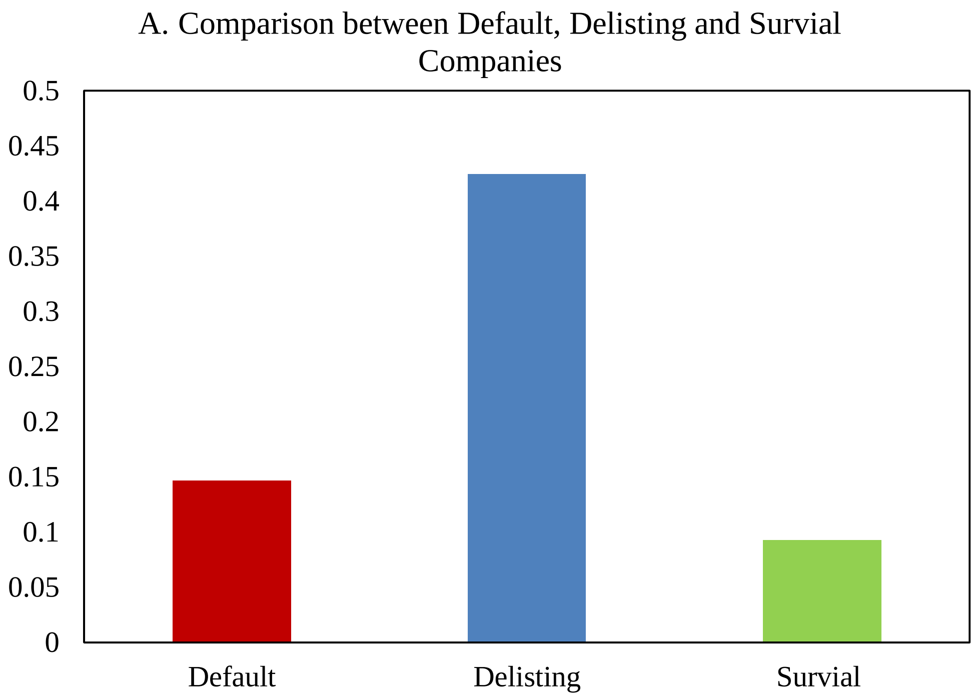
Table 4–3 Estimates for Benchmark Hazard Model

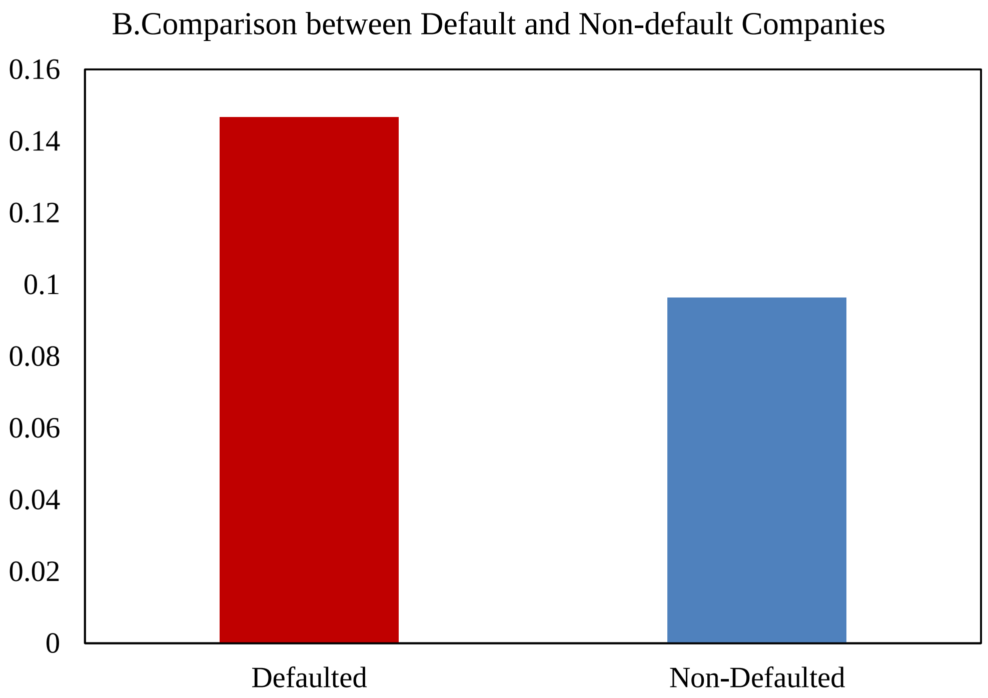
Variables	Model 1	Hazard Ratio
dtd	-0.5082 *** (0.1128)	0.602
cash	-0.7675 *** (0.1347)	0.464
netincome	-0.2170 ** (0.1020)	0.805
m2b	0.1652 *** (0.0466)	1.18
sigma	-0.2964*** (0.1095)	0.743
size	-0.2679 *** (0.0701)	0.765
Industry Fixed Effects	No	
Random Effects	No	
Number of Observations	77,001	
-2 LOG L	2607.692	
AIC	2619.692	
SBC	2639.512	

Note: The definitions of the independent variables are shown in Table 4-2. Standard errors are reported in parentheses.
*Significant at the 10% level;
** Significant at the 5% level;
*** Significant at the 1% level.

Before moving to the model that includes state ownership, a comparison between the ownership structure of default and non-default firms is made in Figure 4-4. The average level of state ownership in companies that have defaulted and that of companies that have never defaulted are compared. Interestingly, it can be seen that companies that enter default have a higher average level of state ownership.

Figure 4-4 State Shares Comparison





Note: In both Panel A and panel B, Y-axis measures the state shares, while X-axis measures corporate type.

Does that mean state ownership is a bad thing? Later results that will be discussed later will show the opposite. Meanwhile, it has been noticed that state-controlled companies have very different characteristics from non-state-controlled companies. As shown in Table 4-4, most of the companies in the sample are not state controlled; only 8% are state-controlled. On average, the level of state ownership in state-controlled companies is nearly 11.5 times that of non-state controlled companies (0.6262/0.0551).

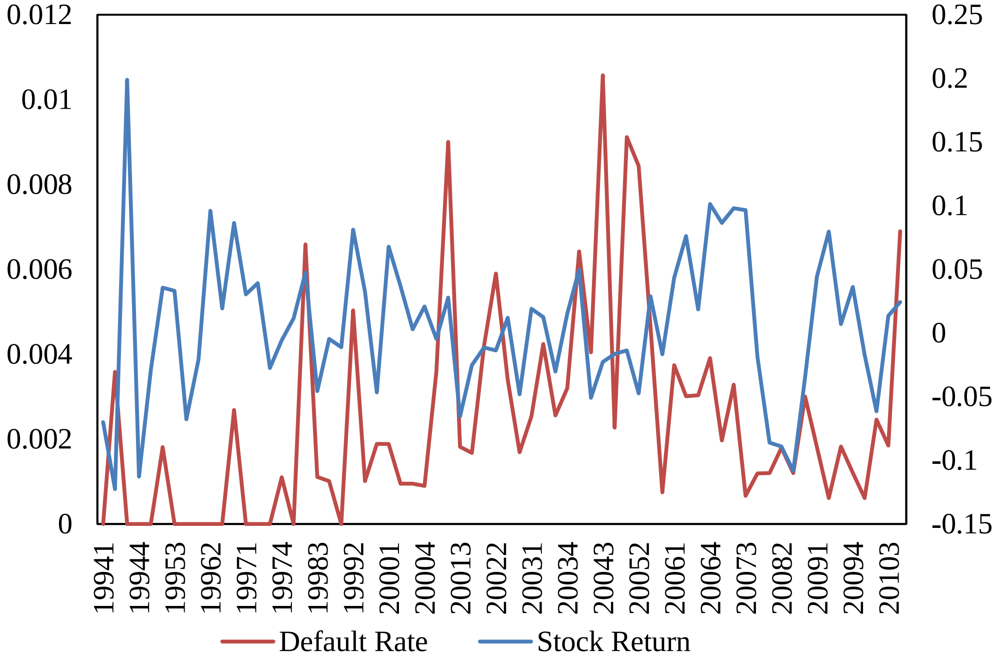
Table 4–4 Comparison between State Controlled and Non-State Controlled Companies

	Obs.	Mean	Std Dev	Minimum	Maximum
State Controlled (state owned shares >50%)	155	0.6262	0.0992	0.5004	0.946
Non-State Controlled (state owned shares =<50%)	1742	0.0551	0.1219	0	0.4994
Total	1897	0.1018	0.1973	0	0.9460

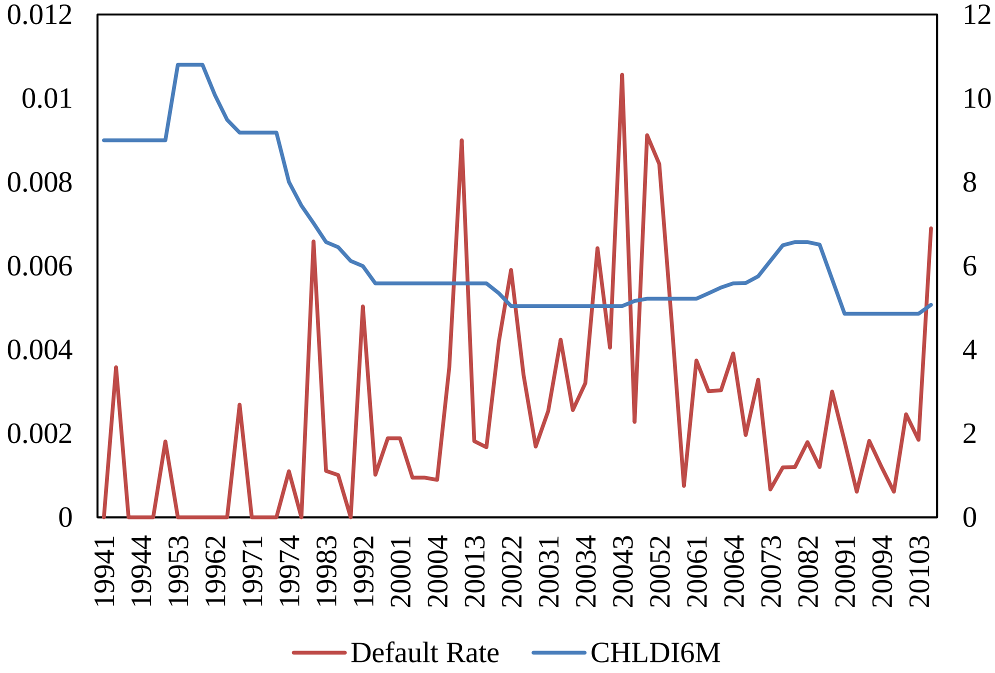
To understand the relation between default probability and macroeconomic variables, each macroeconomic variable eventually used in the model is plotted against the conditional default rate (number of defaults divided by number of companies outstanding) in Figures 4-5. Descriptive statistics about state ownership and macroeconomic variables are reported in Table 4-2.

Figure 4-5 Macro Variables and Default Rate

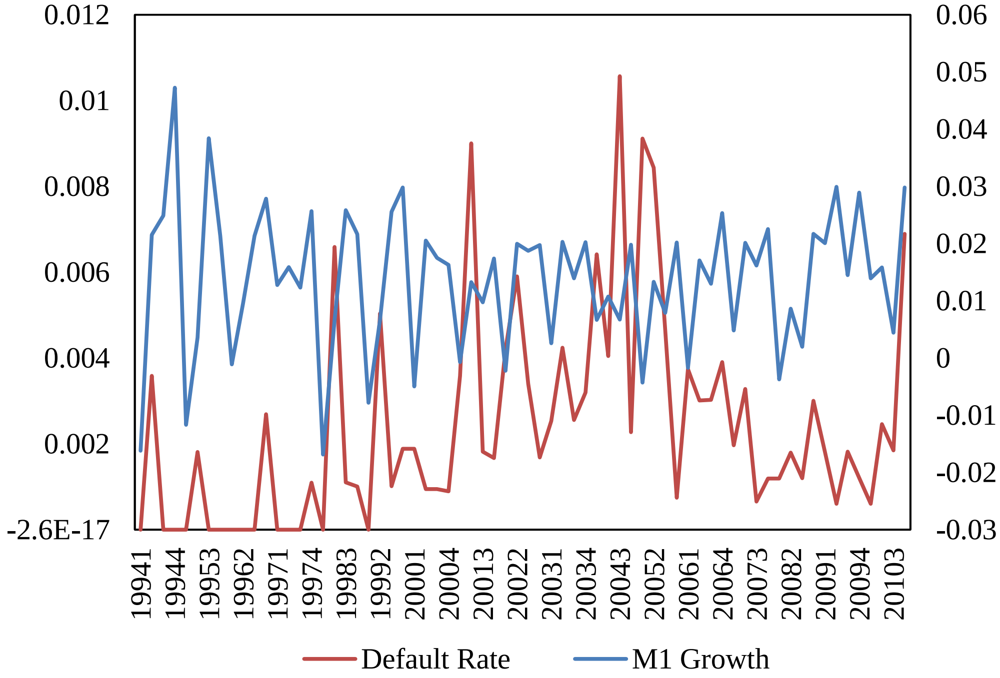
A. Stock Return and Conditional Default Rate



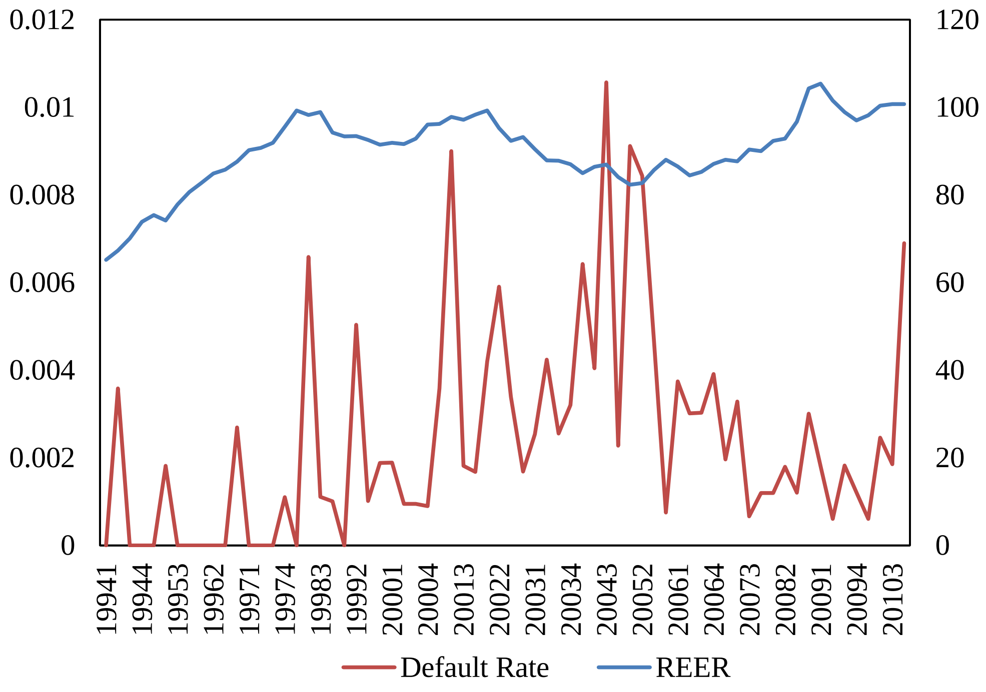
B. Short Term Bank Lending Rate and Conditional Default Rate



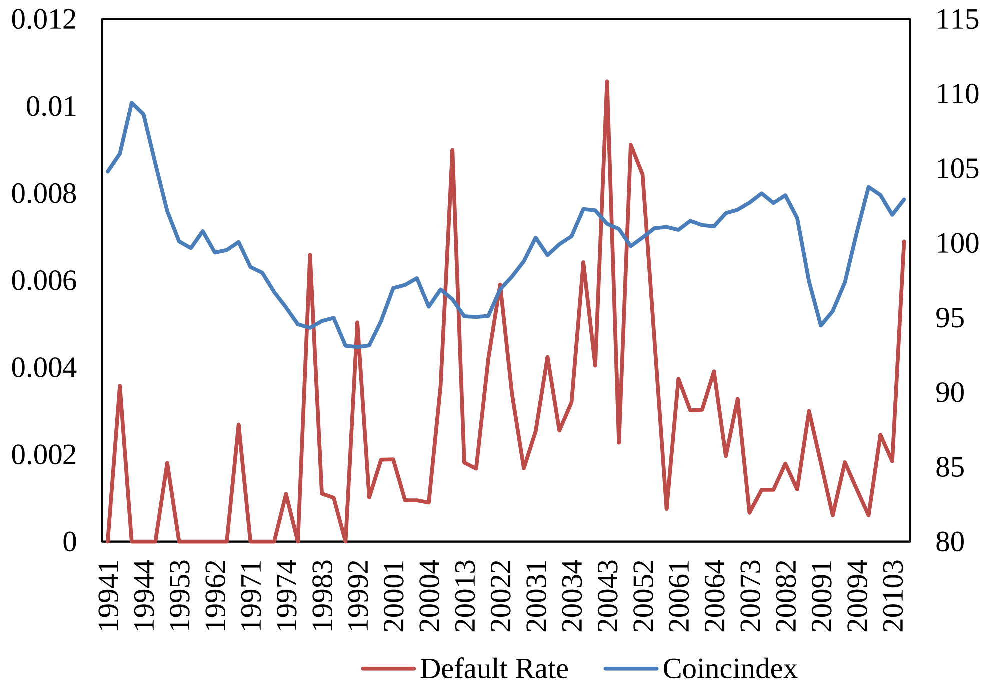
C. Money supply growth and Conditional Default Rate



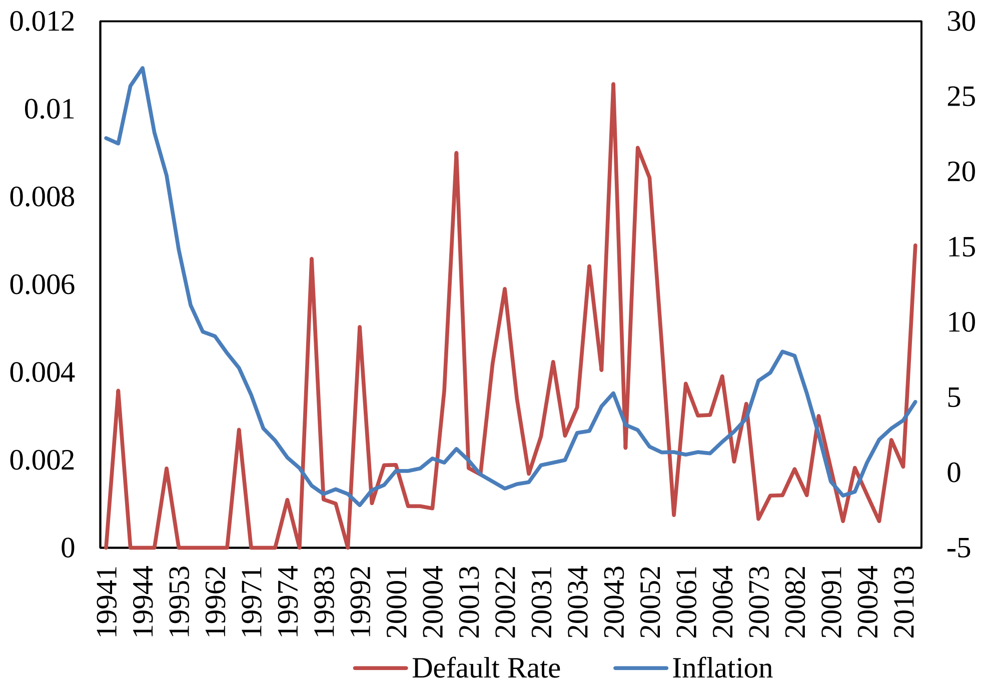
D. Real Effective Exchange Rate and Conditional Default Rate



E. Coincident Index and Conditional Default Rate



F. Inflation and Conditional Default Rate



Note: The definitions of the variables are shown in Table 4-2. The left Y-axis in all panels measures the conditional default rate, while the right Y-axis measure the macroeconomic variables. The X-axis indicates the year.

In Table 4-5, the MLE estimates of the expanded default hazard model, which includes state ownership and macroeconomic variables as well as accounting and market variables, are presented. It can be seen that state ownership has a strong negative impact on default probability. According to the hazard ratio, a fully state-owned company has half the default probability of a completely privately-owned firm. Except for stock market return, all other macroeconomic variables are significantly related to the risk of corporate default. For example, the real effective exchange rate has a negative impact on default probability. This is because the higher the effective exchange rate is, the easier Chinese firms find it to export their goods. Money supply also has a significant impact on default: the higher the growth in money supply, the lower the default probability of Chinese companies. This is possibly because higher money supply leads to easier financing for corporations. However, at the same time, too much money may induce high inflation, which is bad for corporations, as our results show that inflation is positively related to default probability. The coincident index measures the health of the macro economy, and is negatively related to the probability of default. The negative relationship between the short-term lending rate and the probability of default is difficult to explain. Typically, high short-term lending rates mean that the cost of capital is high for corporations, which should have a negative impact on corporate operations and a positive impact on default probability.

Table 4–5 Maximum Likelihood Estimates for Probability of Corporate Default after Including State Ownership and Macroeconomic Variables

Variables	Model 2	Hazard Ratio
dtd	-0.5171*** (0.1134)	0.596
cash	-0.7962 *** (0.1373)	0.451
netincome	-0.1744 * (0.0975)	0.84
m2b	0.1589 *** (0.0461)	1.172
sigma	-0.2338 ** (0.1143)	0.792
size	-0.2115 *** (0.0706)	0.809
State share	-0.5637 *** (0.0984)	0.569
Stock return	0.0994 (0.0954)	1.105
CHLDI6M	-0.5759 ** (0.2422)	0.562
m1gr	-0.3568 ** (0.1694)	0.7
REER	-0.2827** (0.1161)	0.754
coincindex	-0.6608 *** (0.1658)	0.516
inflation	0.8094 *** (0.2698)	2.247
Industry Fixed Effects	No	
Random Effects	No	
Number of Observations	77,001	
-2 LOG L	2548.375	
AIC	2574.375	
SBC	2617.318	

Note: The definitions of the independent variables are shown in Table 4-2. Standard errors are reported in parentheses.
*Significant at the 10% level;
** Significant at the 5% level;
*** Significant at the 1% level.

After including state ownership and macroeconomic variables in the model, none of the accounting and market variables changed their directionality or significance, besides net income which became marginally significant. The other accounting and market variables remained strongly significant. Comparing the AIC and SBC measures in Table 4-3 and Table 4-6, it can be seen that the additional variables improve model fit significantly. In Figure 4-7, the Cumulative Accuracy Profiles (CAP curves) of the benchmark model and the expanded model where state ownership and macroeconomic variables are added are compared. The expanded model dominates the benchmark model in terms of within-sample predicting power. In fact, adding state ownership and macroeconomic variables increases the overall accuracy ratio from 50 percent to 57 percent.

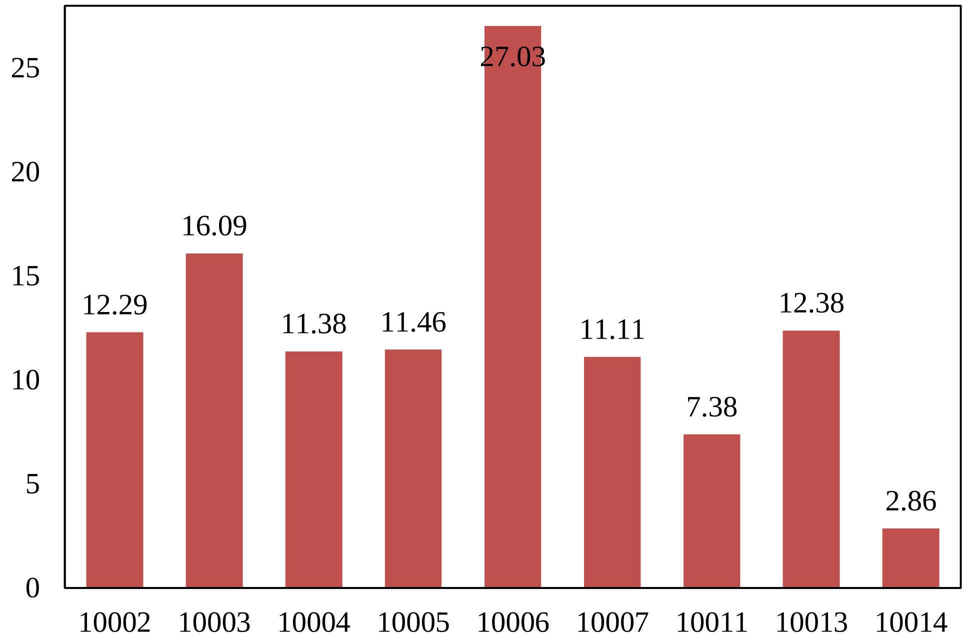
Next, the analysis of the correlation and clustering of corporate defaults within industries is presented. Before proceeding to the results, a comparison of default rates among different industries is presented. From Table 4-6, it can be seen that the consumer and industrial sectors have the highest default rate, following by basic materials and consumer, non-cyclical. The default rate of utilities is the lowest.

Table 4–6 Clustering of Corporate Default within Industries

Industry sector number	Industry Description	Total Obs.	Obs. Of Default	Percentage of Default
10002	Basic Materials	301	37	12.29
10003	Communications	87	14	16.09
10004	Consumer, Cyclical	369	42	11.38
10005	Consumer, Non-cyclical	314	36	11.46
10006	Diversified	37	10	27.03
10007	Energy	45	5	11.11
10011	Industrial	569	42	7.38
10013	Technology	105	13	12.38
10014	Utilities	70	2	2.86
	Total	1,897	201	100

The number of default companies as a percentage of the total number of companies in a particular industry is also presented in Figure 4-6. The highest default rate is found in the diversified sector, followed by the communications sector. Consistent with Table 4-6, the default rate within the utilities industry is still the lowest.

Figure 4-6 Default Rate within Industry



Note: The default rate within industry is calculated as the number of companies entered default as a percentage of the number of companies in a particular industry. The Y-axis measures the default rate, whiele the X-axis measure the industry sector.

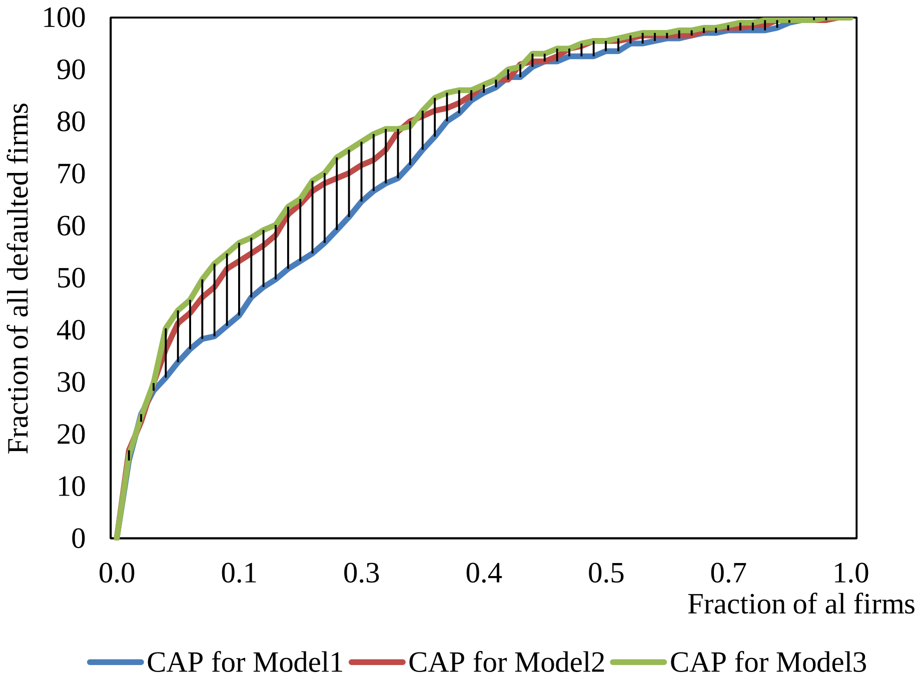
In Table 4-7, the MLE estimates of the fixed-effects model are presented. Here the reference group is the utilities industry. It can be seen that all other eight industries show higher default rates, ceteris paribus. For example, the default risk in the communications, diversified and technology industries are several times that of the utilities industry. All the accounting and market variables, as well as state ownership and macroeconomic variables that are significant in Table 4-5, remain significant. Based on the AIC and SBC measures, it can be seen that adding industry fixed-effects slightly improves the model fit. This is also seen from the CAP curve in Figure 4-7.

Table 4–7 Maximum Likelihood Estimates for Probability of Corporate Default with Industry Fixed Effects

Variables	Model3	Hazard Ratio
dtd	-0.5281 *** (0.1134)	0.59
cash	-0.8385 *** (0.1388)	0.432
netincome	-0.1457 (0.0966)	0.864
m2b	0.1540 *** (0.047)	1.166
sigma	-0.2838 ** (0.1181)	0.753
size	-0.2121 *** (0.0721)	0.809
State share	-0.5417 *** (0.0989)	0.582
Stock return	0.1017 (0.0958)	1.107
CHLDI6M	-0.5677 ** (0.2422)	0.567
m1gr	-0.3554 ** (0.1702)	0.701
REER	-0.2583 ** (0.1171)	0.772
coincindex	-0.6571 *** (0.1672)	0.518
inflation	0.8276 *** (0.2708)	2.288
Industry Sector Dummy Variables		
Materials	1.7987*** (0.7267)	6.042
Communications	2.4178** (0.7597)	11.221
Consumer_Cyclical	1.4593*** (0.7265)	4.303
Consumer_Noncyclical	1.7781** (0.7294)	5.918
Diversified	2.2167*** (0.7763)	9.177
Energy	2.0621** (0.8388)	7.862
Industrial	1.5219** (0.7261)	4.581
Technology	2.3444*** (0.7646)	10.426
Number of Observations	77,001	
-2 LOG L	2520.237	
AIC	2562.237	
SBC	2631.606	

Note: The definitions of the independent variables are shown in Table 4-2. Standard errors are reported in parentheses.
*Significant at the 10% level;
** Significant at the 5% level;
*** Significant at the 1% level.

Figure 4-7 Cumulative Accuracy Profiles



Note: Model 1 is the benchmark hazard model, where only firm-specific accounting and market information are included; Model 2 is the MLE estimates of the expanded default hazard model, which includes state ownership and macroeconomic variables as well as accounting and market variables; Model 3 is the MLE estimates by including the industry fixed-effects.

Finally, the MLE estimates of the random-effects model are presented in Table 4-8. Panel A presents the estimates of all other covariates rather than the industry random effects. It can be seen that they are very consistent with those results in the expanded model (Table 4-5) and those in the fixed-effects model (Table 4-7). Panel B contains our estimate of the random effect parameter θ, the variance of the random effect, and its standard error. Based on the Wald-test results in Panel C, the random effect is significant. Panel D presents the solutions to the random effects of all different industries. The exponentiated estimate is equivalent to the hazard ratio in the covariate estimates. It can be seen that again the default risk of the communications, diversified and technology industries are much higher than that of the utilities industry. However, the differences among the various industries are not as dramatic as those shown in the fixed-effects model. The results of the random effects model suggest that the default risk of Chinese companies is correlated and clustered within industries, and that the correlation is potentially due to common unobservable risk factors.

Table 4–8 Maximum Likelihood Estimates for Probability of Corporate Default with Random Effects

A. Analysis of Maximum Likelihood Estimates		
Variables	Model4	Hazard Ratio
dtd	-0.5264 *** (0.1135)	0.591
cash	-0.8154 *** (0.1383)	0.442
netincome	-0.1558 (0.0969)	0.856
m2b	0.1556 *** (0.0467)	1.168
sigma	-0.2673 ** (0.1169)	0.765
size	-0.2121 *** (0.0721)	0.805
State share	-0.5522 *** (0.0989)	0.576
Stock return	0.1009 (0.0957)	1.106
CHLDI6M	-0.5666** (0.2422)	0.567
m1gr	-0.3566 ** (0.1700)	0.7
REER	-0.2677 ** (0.1167)	0.765
coincindex	-0.6620 *** (0.1669)	0.516
inflation	0.8223 *** (0.2706)	2.276

B. Covariance Parameter Estimates		
Covariance Parameter	REML Estimate	Standard Error
industry_sector	0.1898	0.1301

C. Type 3 Tests		
Effect	Wald Chi-Square	Adjusted Pr > ChiSq
dtd	21.5173	<.0001
cash	34.7652	<.0001
netincome	2.5836	0.1077
m2b	11.0756	0.0009
sigma	5.2318	0.0221
size	9.0784	0.0026
State share	31.2054	<.0001
Stock return	1.1118	0.2917
CHLDI6M	5.4714	0.0193
m1gr	4.4005	0.0359
REER	5.257	0.0218
coincindex	15.7396	<.0001
inflation	9.2326	0.0024
industry_sector_num	18.0078	0.005

D. Solution for Random Effects							
Industry Sector NO.	Estimate	Standard Error	95% CL for Effect		Exponentiated Estimate	95% CL for Exponentiated	
10002	0.0169	0.2159	-0.4064	0.4401	1.017	0.666	1.553
10003	0.4426	0.271	-0.0885	0.9737	1.557	0.915	2.648
10004	-0.2903	0.2102	-0.7023	0.1217	0.748	0.495	1.129
10005	-0.0074	0.2169	-0.4325	0.4177	0.993	0.649	1.519
10006	0.2752	0.2898	-0.2928	0.8431	1.317	0.746	2.324
10007	0.1312	0.3329	-0.5213	0.7837	1.14	0.594	2.19
10011	-0.2345	0.2099	-0.6459	0.177	0.791	0.524	1.194
10013	0.3805	0.2752	-0.1588	0.9199	1.463	0.853	2.509
10014	-0.7142	0.3152	-1.3319	-0.0965	0.49	0.264	0.908
Number of Observations					77,001		
SBC					2631.606		

Note: The definitions of the independent variables are shown in Table 4-2. Standard errors are reported in parentheses.
*Significant at the 10% level;
** Significant at the 5% level;
*** Significant at the 1% level.

4.5 Conclusion

In the context of the United States, accounting variables are good indications of the profitability and health of corporations and thus are found to be predictive of corporate bankruptcy. Meanwhile, market variables such as market price and volatility contain full forward-looking information of a firm's performance and also help predict the likelihood that firms will declare bankruptcy. In fact, Chava and Jarrow (2004) demonstrated that market variables were very informative of bankruptcy and that accounting variables added little predictive power when market variables were included in the bankruptcy model.

However, when applied to Chinese companies, researchers need to worry about the information quality of the accounting and market variables available to them, as the accounting information of Chinese companies is subject to manipulation while the market price and volatility of Chinese stocks are not very informative due to market inefficiency. Therefore, in this chapter, the usefulness of information beyond accounting and market variables in predicting corporate default in China is explored.

The results indicate that accounting and market variables do provide useful information about corporate default. However, other variables such as state ownership, the real effective exchange rate, money supply, short-term lending rate, coincident index and inflation provide additional information. They help increase the predictive power of accounting and market variables. The results also indicate that the default risk of Chinese companies is correlated and clustered within industry. Certain industries such as communications and technology tend to have a higher probability of default risk than others. The results of this essay provide several investment and policy implications. The model built in this chapter will be a tool for risk assessment and management. Future research should try to understand why state ownership helps to reduce corporate default risk.

4.6 Appendix: Estimating Distance-to-Default (DTD)

This appendix shows the calculation method of DTD following the Merton (1974) model and explains the numerical scheme employed to calculate distance-to-default. Merton's model assumes that firms are financed by equity and one single pure discount bond with maturity data T a d principal L. the asset value V_t follows geometric Brownian motion:

$$dV_t = \mu V_t dt + \sigma V_t dB_t$$

Due to limited liability, the equity value at maturity is $E_T = max(V_t - L, 0)$. Therefore, the equity value at time t $\leq$ T by the Black-Scholes option pricing formula becomes:

$$E_t = V_t N(d_t) - e^{-r(T-t)} LN(d_t - \sigma\sqrt{T-t}$$

where r is the instantaneous risk free rate, N(.) is the cumulative distribution function for standard normal random variable, and

$$d_t = \frac{ln(V_t/L) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}$$

According to Merton's model, the company's bankruptcy probability at time t is N($-DTD_t$) where DTD_t denotes distance to default and it is

$$DTD_t = \frac{\ln\left(\frac{V_t}{L}\right) + (\mu - \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}$$

Chapter 5 Conclusion

5.1 Summary of Main Findings

The main findings of this study are as follows.

The first essay studies whether heterogeneity in borrowers' time preferences correlates with their decision to default on their mortgage payments. We hypothesize that naïve borrowers with present-biased preference are more likely to select interest-only loans, so that they can enjoy the immediate benefits of homeownership and postpone their mortgage payment costs. Sophisticated borrowers with present-biased preference, on the other hand, are fully aware of their future self-control problems and know their future preferences exactly, even though they may differ from their current preferences. Therefore, they are more likely to choose 30-year adjustable-rate loans. In contrast, borrowers with time-consistent preference tend to choose 30-year fixed-rate loans, which are fully amortizing mortgage loans where the interest rate on the note remains the same through the term of the loan. Using individual loan-level mortgage data principally collected by BlackBox Logic (BBL) and home loan applications and originations data collected by the Home Mortgage Disclosure Act (HMDA), the empirical analysis comprises a logistic regression, where year of origination and termination and property state location are set as fixed effects, to examine the impact of time preferences on mortgage choice and default decisions. Firstly, the default behavior of naïve borrowers who selected interest-only loans relative to those who selected 30 years fixed-rate loans is studied. The results indicate that borrowers with both 5-year interest-only loans and 10-year interest-only loans are more likely to default than dynamically-consistent borrowers who chose 30-year fixed-rate loans. The next study examines the default behavior of borrowers who chose 30 years adjustable-rate loans relative to 30 years fixed-rate loans. The results indicate that the default rate of sophisticated borrower with present-biased preference is higher than borrowers with time-consistent preference. Lastly, the default behavior of borrowers who take up interest-only loans and 30 years adjustable-rate loans is compared to those who selected 30 years fixed-rate loans. The results indicate that present bias is highly correlated with mortgage default, and borrowers who exhibit present-biased preference in their choice of mortgage have a substantially higher probability of default. The association between present bias and mortgage default holds when controlling for other loan characteristics and housing price. Moreover, all of the results hold after using propensity score matching, based on borrowers' characteristics (including income, race, sex) and loan characteristics (e.g. original loan balance, location of property, origination year etc.). These results are therefore the first direct support for the claim that the mortgage default decisions of borrowers is related to their different time preferences.

The second essay studies the risk of mortgage partial prepayments and the process through which mortgage borrowers learn to make partial prepayment decisions in the residential mortgage market in China. The learning dynamics are measured by studying individual borrowers' repeated mortgage partial prepayment behavior. The empirical model is based on the conditional fixed effects multinomial logit model (FEMNL). Using a rich set of individual mortgage loan history data from a leading mortgage lender in China, the empirical results indicate that option theory is not applicable to China, and that other investment opportunities are important in explaining mortgage partial prepayment behavior. Moreover, borrowers' characteristics, such as age, occupation, job position, gender, and income, significantly impact partial prepayment behavior. Thus, these attributes may be used to screen loan applicants and for determining potential high-risk borrowers. Lastly, and most importantly, a borrower's partial prepayment behavior follows the reinforcement learning process. A borrower's partial prepayment decision not only depends upon current stage variables (like other investment opportunities) and borrower characteristics, but also learning through past experience. Borrowers who have made partial prepayments in earlier stages are more likely to make the same decision in the future. In particular, learning from experience increases the probability of partial prepayment by around 26.9 percentage points, and the experience of learning from others increases the probability of partial prepayment by around 1.8 percentage points. Moreover, the results indicate that learning dynamics are not monotonic, and recent experience plays a larger role than older experience in determining the partial prepayment decision.

The third essay studies the value of information besides accounting and market variables in predicting the default of Chinese companies. In the US market, accounting and market variables provide useful information about corporate default. However, the information quality of accounting and market variables is often poorer in developing countries. To predict the default probability of Chinese-listed companies, data from the National University of Singapore (NUS) Risk Management Institute's (RMI) corporate default database is used. Firstly, a standard Cox proportional hazard model on the default probability of Chinese companies using firm-level accounting and market variables is tested. The results indicated that accounting and market variables provide useful information about corporate default. Next, the presence of state ownership and macroeconomic variables, such as short-term bank lending rate, effective exchange rate, growth in money supply, coincident index and inflation, were added into the model. These variables are found to provide additional information on the default of Chinese-listed corporations and significantly increasing the predictive power of accounting and market variables. The firms in the sample were classified into nine categories, following the Bloomberg industry classification. The results indicated that the default risk of Chinese companies was correlated and clustered within industries. Certain industries, such as communications and technology, tended to have higher default risk than others. The model built in this essay will be useful for risk assessment and management.

5.2 Contributions

This study fills the knowledge gap of loan repayment behavior of household and corporate from three aspects: residential mortgage default behavior, residential mortgage partial prepayment behavior and corporate default risk. More specifically:

The first essay contributes to understanding current mortgage default behavior by presenting an alternative theory to explain the origins of the unobserved heterogeneity of mortgage borrowers and how this unobserved heterogeneity affects borrowers' default decisions and behavior. In the literature on mortgage default, many empirical studies have focused on the effects of transaction costs and unobserved heterogeneity on mortgage default. However, there has been no unifying theory to explain the underlying unobserved heterogeneity of borrowers. This essay filled this gap by assuming that this unobserved heterogeneity is based on borrowers' time preferences. In addition, the study presented in this essay also sheds light on the studies of present-biased preferences, which have been extensively addressed both in psychology and behavioral economics. Past research has found that present bias explains job search behavior, and that individual differences in time preference are an important predictor of many life outcomes, including gym contracts, smoking propensity, body-mass index, savings towards retirement, and credit card debt. However, researchers have not studied the effects of present bias on mortgage choice and default, and relatively few papers on mortgage choice and default have distinguished between naïve and sophisticated individuals. In this essay, heterogeneous time preferences among borrowers, as indicated by their mortgage choices, are used to explain the default behavior of present-biased borrowers. Different mortgage types are also used to differentiate naïve and sophisticated borrowers.

The second essay contributes to the current mortgage payment literature by exploring partial prepayments. Despite the existence of many studies on default and prepayment behavior, few studies have paid attention to the risk of partial prepayment and the corresponding behavior of borrowers. In addition to filling in the gap in partial prepayment risk studies in mortgage markets, this essay also contributes to the literature of reinforcement learning by studying the repeat partial prepayment behavior of individual borrowers. While many economic studies have analyzed learning in the laboratory or in the field, only a few papers have measured learning with household-level panel data, because of the challenges of getting such data. This essay fills this gap by studying learning behavior using household-level mortgage payment data.

The third essay contributes to the literature on corporate default by exploring the usefulness of information beyond accounting and market variables in predicting Chinese corporate default. The model built in this study is useful for risk assessment and management, and also contributes to the investment community. In addition, this essay has strong policy implications. For example, it found that the effective exchange rate has a negative relationship with corporate default, which suggests that depreciating the Chinese RMB would not only reduce the profitability of many Chinese companies but also significantly increase the default risk of corporations. Money supply, bank lending rate, and inflation also significantly impact corporate default, suggesting that policy makers should take into consideration the impact of monetary policy on corporate default.

5.3 Limitations and Future Works

No research is without limitations and this study is no exception. This section highlights the limitations of each of the essays, together with recommendations for further research.

While the first essay only distinguished between sophisticate and fully naïve individuals, O'Donoghue and Rabin (1999a) differentiated between naïve and partially naïve agents. Partially naïve agents believed their future selves would act with some short-term discount factor, $\tilde{\beta} > \beta$. In the case of completely naïve individuals, $\tilde{\beta} = \beta = 1$. In the future, a survey of individuals could be done to identify each person's time preference so that they can be assigned into one of three groups: sophisticates, full naïf and partial naïf. Moreover, future research could focus on the more difficult problem of exploring the theoretically-proposed causal link between present bias and mortgage default.

In emerging markets, and especially in China, the availability of data is a challenge for academic research. The residential mortgage data used in the second essay is obtained from a leading lender from China. When issuing a mortgage loan, different lenders have different criteria for assessing the risk of borrowers. With the absence of lender variation, hidden risks affecting the probability of mortgage payments may be neglected. Future research should use more comprehensive data, which can be obtained from different banks in China, and capture the risk caused by lenders' variation.

In the third essay, a generally negative relationship between state ownership and the probability of corporate default is shown. However, there is no detailed exploration of how this occurs and why state ownership should reduce the risk of corporate default. Future researchers should try to understand why state ownership can help to reduce corporate default risk. In addition, as the random effects model suggests, the intra-industry default correlation is indeed due to common latent factors, and source of default risk cannot be diversified away. The time-series characteristics of the frailty factor should be further studied, for example by following Duffie et al. (2009), to better understand portfolio default risk.

BIBLIOGRAPHY

Aaronson, Daniel, Sumit Agarwal, and Eric French. 2012. "The Spending and Debt Response to Minimum Wage Hikes." *American Economic Review* 102(7): 3111-3139.

Abrahams, Steven W. 1997. "The New View in Mortgage Prepayments: Insight from Analysis at the Loan-by-Loan Level." *Journal of Fixed Income* 7(1): 8–21.

Agarwal, Sumit, John C. Driscoll, and David I. Laibson. 2013. "Optimal Mortgage Refinancing: A Closed-Form Solution." *Journal of Money, Credit, and Banking* 45(4): 591-622.

Agarwal, Sumit, John C. Driscoll, Xavier Gabaix, and David I. Laibson. 2008. "Learning in the Credit Card Market." *NBER Working Paper* No.13822.

Agarwal, Sumit, Souphala Chomsisengphet, Chunlin Liu, and Nicholas S. Souleles. 2006. "Do Consumers Choose the Right Credit Contracts?" *Federal Reserve Bank of Chicago Working Paper* No. wp-06-11.

Agarwal, Sumit, Yan Chang, and Abdullah Yavas. 2012. "Adverse Selection in Mortgage Securitization." *Journal of Financial Economics* 105(3): 640-660.

Agarwal, Vineet and Richard Taffler. 2008. "Comparing the Performance of Market-based and Accounting-based Bankruptcy Prediction Models." *Journal of Banking and Finance* 32(8): 1541-1551.

Altman, Edward I. 1968. "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy." *Journal of Finance* 23(4): 589-609.

Altman, Edward I. 2012. "Predicting Financial Distress of Companies: Revisiting the Z-Score and Zeta Models." in *Handbook of Research Methods and Applications in Empirical Finance*, edited by A. R. Bell, C. Brooks, and M. Prokopczuk. Cheltenham: Edward Elgar Publishing.

Altman, Edward I., Robert G. Haldeman, and Paul Narayanan. 1977. "Zeta Analysis: A New Model to Identify Bankruptcy Risk of Corporations." *Journal of Banking and Finance* 1(1): 29-54.

An Xudong, Yongheng Deng, and Stuart A. Gabriel. 2011. "Asymmetric Information, Adverse Selection, and the Pricing of CMBS." *Journal of Financial Economics* 100(2): 304-325.

An, Xudong, Yongheng Deng, Joseph B. Nichols, and Anthony B. Sanders. 2013. "Local Traits and Securitized Commercial Mortgage Default." *Journal of Real Estate Finance and Economics* 47(4): 787-813.

Archer, Wayne R., David C. Ling, and Gary A. McGill. 1996. "The Effect of Income and Collateral Constrains on Residential Mortgage Terminations." *Regional Science and Urban Economics* 26(3-4): 235-261.

Archer, Wayne R., David C. Ling, and Gary A. McGill. 1997. "Demographic Versus Option-Driven Mortgage Terminations." *Journal of Housing Economics* 6(2): 137-163.

Archer, Wayne R., Peter J. Elmer, David M. Harrison, and David C. Ling. 2002. "Determinants of Multifamily Mortgage Default." *Real Estate Economics* 30(3): 445-473.

Arthur, W. Brain. 1991. "Designing Economic Agents that Act Like Human Agents: A Behavioral Approach to Bounded Rationality." *American Economic Review* 81(2): 353-359.

Arthur, W. Brain. 1993. "On Designing Economic Agents that Behave Like Human Agents." *Journal of Evolutionary Economics* 3(1): 1-22.

Asay, Michael R. 1978. "Rational Mortgage Pricing." PhD dissertation, University of California, Los Angeles.

Axel, Borsch S. 1990. "Panel Data Analysis of Housing Choices." *Regional Science and Urban Economics* 20(1): 65-82.

Bajari, Patrick, Chenghua S. Chu, and Mingjun Park. 2008. "An Empirical Model of Subprime Mortgage Default from 2000 to 2007." *NBER Working Paper* No.14625.

Bangia, Anil, Francis X. Diebold, Andre Kronimus, Christian Schagen, and Til Schuermann. 2002. "Rating Migration and the Business Cycle, with Application to Credit Portfolio Stress Testing." *Journal of Banking and Finance* 26(2-3): 445-474.

Beaver, William H. 1966. "Financial Ratios as Predictors of Failure." *Journal of Accounting Research* 4(3): 71-111.

Beaver, William H., Maureen F. McNichols, and Jung-Wu Rhie. 2005. “Have Financial Statements Become Less Informative? Evidence from the Ability of Financial Ratios to Predict Bankruptcy.” *Review of Accounting Studies* 10(1): 93-122.

Becker, Gary S. 1965. “A Theory of the Allocation of Time.” *The Economic Journal* 75(299): 493-517.

Becker, Gary S. 1975. *Human Capital*. 3rd Edition. Chicago, IL: University of Chicago Press.

Bharath, Sreedhar T. and Tyler Shumway. 2008. “Forecasting Default with the Merton Distance to Default Mode.” *Review of Financial Studies* 21(3): 1339-1369.

Bhutta, Neil, Jane Dokko, and Hui Shan. 2010. “The Depth of Negative Equity and Mortgage Default Decisions.” *Finance and Economics Discussion Series* No.2010-35.

Black, Fischer S. and John C. Cox. 1976. “Valuing Corporate Securities: Some Effects of Bond Indenture Provisions.” *Journal of Finance* 31(2): 351-367.

Black, Fischer S. and Myron S. Scholes. 1973. “The Pricing of Options and Corporate Liabilities.” *Journal of Political Economy* 81(3):637-654.

Blum, Marc. 1974. “Failing Company Discriminant Analysis.” *Journal of Accounting Research* 12(1): 1-25.

Brennan, Michael J. and Eduardo S. Schwartz. 1985. “Determinants of GNMA Mortgage Prices.” *Real Estate Economics* 13(3): 209-228.

Buser. Stephen A. and Patric H. Hendershott. 1984. “The Pricing of Default-Free Fixed Rate Mortgages.” *Housing Finance Review* 3(4): 405-429.

Bush, Robert R. and Frederick Mosteller. 1955. *Stochastic Models for Learning*. New York: John Wiliey and Son.

Campbell, John Y., Jens Hilscher, and Jan Szilagyi. 2008. “In Search of Distress Risk.” *Journal of Finance* 63(6): 2899-2939.

Carrillo, Juan D. 1999. ‘‘Self-Control, Moderate Consumption, and Craving.’’ *CEPR Discussion Paper* No.2017.

Carroll, Gabriel D., James J. Choi, David I. Laibson, Brigitte C. Madrian, and Andrew Metrick. 2009. "Optimal Default and Active Decisions." *Quarterly Journal of Economics* 124(4): 1639-1674.

Chabris, Christopher F., David I. Laibson, and Jonathon P. Schuldt. 2008. "Intertemporal choice." in *The New Palgrave Dictionary of Economics* (Second Edition), edited by S. N. Durlauf and L. E. Blume. London: Palgrave Macmillan.

Chamberlain, Gary. 1980. "Analysis of Covariance with Qualitative Data." *Review of Economic Studies* 47(1): 225-238.

Charalambakis, Evangelos C., Susanne K. Espenlauby, and Ian Garrett. 2009. "Assessing the Probability of Financial Distress of UK Firms." *Working Paper*.

Chava, Sudheer and Robert A. Jarrow. 2004. "Bankruptcy Prediction with Industry Effects." *Review of Finance* 8(4): 537-569.

Chinloy, Peter. 1993. "Elective Mortgage Prepayment: Termination and Curtailment." *Real Estate Economics* 21(3): 313-332.

Clapp, John M., Yongheng Deng, and Xudong An. 2006. "Unobserved Heterogeneity in Models of Competing Mortgage Termination Risks." *Real Estate Economics* 34(2): 243-273.

Clauretie, Terrence M. 1987. "The Impact of Interstate Foreclosure Cost Differences and the Value of Mortgages on Default Rates." *Real Estate Economics* 15(3): 152-167

Courtemanche, Charles and Art Carden. 2011. "Supersizing Supercenters? The Impact of Walmart Supercenters on Body Mass Index and Obesity." *Journal of Urban Economics* 69(2): 165–181.

Cox, John C., Jonathan E. Ingersoll, Jr, and Stephen A. Ross. 1985. "A Theory of the Term Structure of Interest Rates." *Econometrica* 53(2): 385-408.

Crosbie, Peter and Jeff Bohn. 2002. "Modeling Default Risk." KMV.

Cross, John G. 1973. "A Stochastic Learning Model of Economic Behavior." *Quartertly Journal of Economics* 87(2): 239-266.

Cunningham, Donald F. and Patric H. Hendershott. 1984. "Pricing FHA Mortgage Default Insurance." *Housing Finance Review* 3(4): 373-392.

DellaVigna, Stefano and M. Daniele Paserman. 2005. “Job Search and Impatience.” *Journal of Labor Economics* 23(3): 527-588.

DellaVigna, Stefano and Ulrike Malmendier. 2006. “Paying Not to Go to the Gym.” *American Economic Review* 96(3): 694-719.

Demyanyk, Yuliya and Otto V. Hemert. 2011. “Understanding the Subprime Mortgage Crisis.” *Review of Financial Studies* 24(6): 1848-1880.

Deng, Yongheng. 1997. “Mortgage Termination: An Empirical Hazard Model with Stochastic Term Structure.” *Journal of Real Estate Finance and Economics* 14(3): 309-331.

Deng, Yongheng and John M. Quigley. 2012. “Woodhead Behavior and the Pricing of Residential Mortgages.” *NUS Institute of Real Estate Studies Working Paper Series* No.2012-025.

Deng, Yongheng and Peng Liu. 2009. “Mortgage Prepayment and Default Behavior with Embedded Forward Contract Risks in China’s Housing Market.” *Journal of Real Estate Finance and Economics* 38(3): 214-240.

Deng, Yongheng, Andrey D. Pavlov, and Lihong Yang. 2005. “Spatial Heterogeneity in Mortgage Terminations by Refinance, Sale and Default.” *Real Estate Economics* 33(4): 739-764.

Deng, Yongheng, Della Zheng, and Changfeng Ling. 2005. “An Early Assessment of Residential Mortgage Performance in China.” *Journal of Real Estate Finance and Economics* 31(2): 117-136.

Deng, Yongheng, John M. Quigley, and Robert Van Order. 1996. “Mortgage Default and Low Down-payment Loans: The Cost of Public Subsidy.” *Regional Science and Urban Economics* 26(3-4): 263-285.

Deng, Yongheng, John M. Quigley, and Robert Van Order. 2000. “Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options.” *Econometrica* 68(2): 275-307.

Deng, Yongheng, Randall Mock, Jing Wu, and Bernard Yeung. 2011. “Monetary and Fiscal Stimuli, Ownership Structure, and China’s Housing Market.” forthcoming in *Review of Finance*.

Duan, Jin-Chuan, Jie Sun, and Tao Wang. 2012. “Multiperiod Corporate Default Prediction – A Forward Intensity Approach.” *Journal of Econometrics* 170(1): 191-209.

Duffie, Darrell, Andereas Eckner, Guillaume Horel, and Leandro Saita. 2009. "Frailty Correlated Default." *Journal of Finance* 64(5): 2089-2123.

Dunn, Kenneth B. and John J. McConnell. 1981a. "A Comparison of Alternative Models for Pricing GNMA Mortgage-Backed Securities." *Journal of Finance* 36(2): 471-483.

Dunn, Kenneth B. and John J. MoConnell. 1981b. "Valuation of GNMA Mortgage-Backed Securities." *Journal of Finance* 36(3): 599-617.

Erev Ido and Alvin E. Roth. 1998. "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria." *American Economic Review* 88(4): 848–881.

Fischer, Carolyn. 2001. "Read This Paper Even Later: Procrastination with Time-Inconsistent Preferences." *Journal of Economic Behavior and Organization* 46(3): 249-269.

Foster, Chester and Robert Van Order. 1984. "An Option-based Model of Mortgage Default." *Housing Finance Review* 3(4): 351-372.

Foster, Chester and Robert Van Order. 1985. "FHA Terminations: A Prelude to Rational Mortgage Pricing." *Real Estate Economics* 13(3): 292-316.

Frederick, Shane, George Loewenstein, and Ted O'Donoghue. 2002. "Time Discounting and Time Preference: A Critical Review." *Journal of Economic Literature* 40(2): 351–401.

Gentry, James A., Paul Newbold, and David T. Whitford. 1985. "Classifying Bankrupt Firms with Funds Flow Components." *Journal of Accounting Research* 23(1): 146-161.

Gentry, James A., Paul Newbold, and David T. Whitford. 1985. "Predicting Bankruptcy: If Cashflow's Not the Bottom Line, What Is?" *Financial Analysts Journal* 41(5): 47-56.

Green, Jerry R. and John B. Shoven. 1986. "The Effect of Interest Rates on Mortgage Prepayment." *Journal of Money, Credit and Banking* 18(1): 41-50.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2013. "The Determinants of Attitudes toward Strategic Default on Mortgages." *Journal of Finance* 68(4): 1473-1515.

Gul, Faruk and Wolfgang Pesendorfer. 2007. "Harmful Addiction." *Review of Economic Studies* 74(1): 147-172.

Hall, Arden. 2000. "Controlling for Burnout in Estimating Mortgage Prepayment Models." *Journal of Housing Economics* 9(4): 215-232.

Harding, John P. 1997. "Estimating Borrower Mobility from Observed Prepayments." *Real Estate Economics* 25(3): 347-371.

Haselhuhn, Michael P., Devin G. Pope, Maurice E. Schweitzer, and Peter Fishman. 2012. "The Impact of Personal Experience on Behavior: Evidence from Video-Rental Fines." *Management Science* 58(1): 52-61.

Hayre, Lakhbir S. and Kenneth Lauterbach. 1991. "Partial and Full Prepayments and the Modeling of Mortgage Cash Flows." *Journal of Fixed Income* 1(2): 37-42.

Helfert, Erich. 1982. *Techniques in Financial Analysis*. 5th Edition. Homewood, IL: Richard D. Irwin.

Hendershott, Patric H. and William R. Schultz. 1993. "Equity and Nonequity Determinants of FHA Single-Family Mortgage Foreclosures in the 1980s." *Real Estate Economics* 21(4): 405-430.

Hillegeist, Stephen A., Elizabeth K. Keating, Donald P. Cram, and Kyle G. Lundstedt. 2004. "Assessing the Probability of Bankruptcy." *Review of Accounting Studies* 9(1): 5-34.

Ho, Teck-Hua and Juin-Kuan Chong. 2003. "A Parsimonious Model of Stockkeeping-Unit Choice." *Journal of Marketing Research* 40(3): 351-365.

Kalbfleisch, John D. and Ross L. Prentice. 1980. *The Statistical Analysis of Failure Time Data*. New York: John Wiley & Sons.

Kau, James B. and Donald C. Keenan. 1996. "An Option-Theoretic Model of Catastrophes Applied to Mortgage Insurance." *Journal of Risk and Insurance* 63(4): 639-656.

Kau, James B. and V. Carlos Slawson, Jr. 2002. "Frictions, Heterogeneity and Optimality in Mortgage Modeling." *Journal of Real Estate Finance and Economics* 24(3): 239-260.

Kau, James B., Donald C. Keenan, and Taewon Kim. 1993. "Transaction Costs, Suboptimal Termination, and Default Probabilities." *Real Estate Economics* 21(3): 247-263.

Kau, James B., Donald C. Keenan, and Taewon Kim. 1994. "Default Probabilities for Mortgages." *Journal of Urban Economics* 35(3): 278-296.

Kau, James B., Donald C. Keenan, Walter J. Mueller, and James F. Epperson. 1992. "A Generalized Valuation Model for Fixed-Rate Residential Mortgages." *Journal of Money, Credit, and Banking* 24(3): 279-299.

Kau, James B., Donald C. Keenan, Walter J. Mueller, and James F. Epperson. 1995. "The Valuation at Origination of Fixed-Rate Mortgages with Default and Prepayment." *Journal of Real Estate Finance and Economics* 11(1): 5-36.

Keys, Benjamin J., Amit Seru, and Vikrant Vig. 2012. "Lender Screening and the Role of Securitization: Evidence from Prime and Subprime Mortgage Markets." *Review of Financial Studies* 25(7): 2071-2108.

Keys, Benjamin J., Tanmoy K. Mukherjee, Amit Seru, and Vikrant Vig. 2010. "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans." *Quarterly Journal of Economics* 125(1): 307-362.

Kiff, John and Paul S. Mills. 2007. "Money for Nothing and Checks for Free: Recent Developments in U.S. Subprime Mortgage Markets." *International Monetary Fund Working Paper* No.07-188.

Krusell, Per, Burhanettin Kuruscu, and Anthony A. Smith, Jr. 2010. "Temptation and Taxation." *Econometrica* 78(6): 2063-2084.

Kuchler, Theresa. 2013. "Sticking to Your Plan: Hyperbolic Discounting and Credit Card Debt Paydown." *Stanford Institute for Economic Policy Research Discussion Paper* No.12-025.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert W. Vishny. 1997. "Legal Determinants of External Finance." *Journal of Finance* 52(3): 1131-1150.

Laibson, David I. 1994. "Essays in Hyperbolic Discounting." PhD dissertation, Massachusetts Institute of Technology.

Laibson, David I. 1997. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics* 112(2): 443-447.

Laibson, David I. 1998. "Life-Cycle Consumption and Hyperbolic Discount Functions." *European Economic Review* 42(3-5): 861-871.

Laibson, David I., Andrea Repetto, and Jeremy Tobacman. 2007. "Estimating Discount Functions with Consumption Choices over the Lifecycle." *NBER Working Paper* No.13314.

Lau, Amy Hing-Ling. 1987. "A Five-State Financial Distress Prediction Model." *Journal of Accounting Research* 25(1): 127–138.

Lekkas, Vassilis, John M. Quigley, and Robert Van Order. 1993. "Loan Loss Severity and Optimal Mortgage Default." *Real Estate Economics* 21(4): 353–371.

Lennox, Clive. 1999. "Identifying Failing Companies: A Re-evaluation of the Logit, Probit and DA Approaches." *Journal of Economics and Business* 51(4): 347-364.

Lin, Yifu and Guofu Tan. 1999. "Policy Burdens, Accountability, and the Soft Budget Constraint." *American Economic Review* 89(2): 426-431.

Lin, Yifu and Zhiyun Li. 2008. "Policy Burden, Privatization and Soft Budget Constraint." *Journal of Comparative Economics* 36(1): 90-102.

Meier, Stephan and Charles Sprenger. 2010. "Present-Biased Preferences and Credit Card Borrowing." *American Economic Journal: Applied Economics* 2(1): 193-210.

Menash, Yaw M. 1984. "An Examination of the Stationarity of Multivariate Bankruptcy Prediction Models: A Methodological Study." *Journal of Accounting Research* 22(1): 380-395.

Merton, Robert C. 1973. "Theory of Rational Option Pricing." *The Bell Journal of Economics and Management Science* 4(1): 141-183.

Merton, Robert C. 1974. "On the Pricing of Corporate Debt: the Risk Structure of Interest Rates." *Journal of Finance* 29(2): 28-30.

Mian, Atif and Amir Sufi. 2009. "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis." *Quarterly Journal of Economics* 124(4): 1449-1496.

Miravete, Eugenio J. 2003. "Choosing the Wrong Calling Plan? Ignorance and Learning." *American Economic Review* 93(1): 297-310.

Moody's. 2006. "Moody's Credit Rating Prediction Model." *Moody's Special Comment*.

Morck, Randall, Bernard Yeung, and Wayne Yu. 2000. "The Information Content of Stock Markets: Why do Emerging Markets have Synchronous Stock Price Movements?" *Journal of Financial Economics* 58(1): 215-260.

Nickell, Pamela, William Perraudin, and Simone Varotto. 2000. “Stability of Rating Transitions.” *Journal of Banking and Finance* 24(1-2): 203-227.

O'Donoghue, Ted and Matthew J. Rabin. 1999a. “Doing It Now or Later.” *American Economic Review* 89(1): 103-124.

O'Donoghue, Ted and Matthew J. Rabin. 1999b. “Addiction and Self Control.” in *Addiction: Entries and Exits*, edited by J. Elster. New York: Russell Sage Foundation.

O'Donoghue, Ted and Matthew J. Rabin. 1999c. “Procrastination in Preparing for Retirement.” in *Behavioral Dimensions of Retirement Economics*, edited by H. Aaron. New York: Brookings Institution Press and Russell Sage Foundation.

O'Donoghue, Ted and Matthew J. Rabin. 2001. “Choice and Procrastination.” *Quarterly Journal of Economics* 116(1): 121–160.

O'Donoghue, Ted and Matthew J. Rabin. 2002. “Addiction and Present-Biased Preferences.” *Center for Analytical Economics Working Paper* No.02-10.

O'Donoghue, Ted and Matthew Rabin. 2008. “Procrastination on Long-term Projects.” *Journal of Economic Behavior and Organization* 66(2): 161-175.

Ohlson, James A. 1980. “Financial Ratios and the Probabilistic Prediction of Bankruptcy.” *Journal of Accounting Research* 18(1): 109-131.

Phelps, Edmund S. and Robert A. Pollak. 1968. “On Second-Best National Savings and Game-Equilibrium Growth.” *Review of Economic Studies* 35(2): 185-199.

Prelec, Drazen and George Loewenstein. 1998. “The Red and the Black: Mental Accounting of Savings and Debt.” *Marketing Science* 17(1): 4-28.

Quercia, Roberto G. and Michael A. Stegman. 1992. “Residential Mortgage Default: A Review of the Literature.” *Journal of Housing Research* 3(2): 341–379.

Quigley, John M. and Robert Van Order. 1991. “Defaults on Mortgage Obligations and Capital Requirements for U.S. Savings Institutions: A Policy Perspective.” *Journal of Public Economics* 44(3): 353-369.

Quigley, John M. and Robert Van Order. 1995. “Explicit Tests of Contingent Claims Models of Mortgage Default.” *Journal of Real Estate Finance and Economics* 1(2): 99-117.

Rasch, Grorge. 1960. *Probabilistic Models for Some Intelligence and Attainment Tests.* Copenhagen: The Danish Institute for Educational Research.

Reimers, Stian, Elizabeth A. Maylor, Neil Stewart, and Nick Chater. 2009. "Associations between a One-Shot Delay Discounting Measure and Age, Income, Education and Real-World Impulsive Behavior." *Personality and Individual Differences* 47(8): 973-978.

Reisz, Alexander S. and Claudia Perlich. 2007. "A Market-Based Framework for Bankruptcy Prediction." *Journal of Financial Stability* 3(2): 85–131.

Richard, Scott F. and Richard Roll. 1989. "Prepayments on Fixed-Rate Mortgage-Backed Securities." *Journal of Portfolio Management* 15(3): 73-82.

Roth, Alvin E. and Ido Erev. 1995. "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term." *Games and Economic Behavior* 8(1): 164–212.

Saunders, Anthony and Linda Allen. 2002. *Credit Risk Measurement: New Approaches to Value at Risk and Other Paradigms*. 2nd Edition. New York: John Wiley and Sons.

Schwartz, Eduardo S. and Walter N. Torous. 1989. "Prepayment and the Valuation of Mortgage-Backed Securities." *Journal of Finance* 44(2): 375-392.

Shleifer, Andrei, and Robert W. Vishny. 1994. "Politicians and Firms." *Quarterly Journal of Economics* 109(4): 995-1025.

Shumway, Tyler. 2001. "Forecasting Bankruptcy More Accurately: A Simple Hazard Model." *Journal of Business* 74(1): 101-124.

Smith, Patricia K., Barry Bogin, and David Bishai. 2005. 'Are Time Preference and Body Mass Index Associated? Evidence from the National Longitudinal Survey of Youth." *Economics and Human Biology* 3(2): 259–270.

Stanton, Richard. 1995. "Rational Prepayment and the Valuation of Mortgage-Backed Securities." *Review of Financial Studies* 8(3): 677-708.

Stanton, Richard. 1996. "Unobservable Heterogeneity and Rational Learning: Pool Specific vs. Generic Mortgage-Backed Security Prices." *Journal of Real Estate Finance and Economics* 12(3): 243-263.

Taffler, Richard J. 1983. "The Assessment of Company Solvency and Performance Using a Statistical Model." *Accounting and Business Research*. 13(52): 295-308.

Therneau, Terry M. and Patricia M. Grambsch. 2000. *Modeling Survival Data: Extending the Cox Model*. New York: Springer-Verlag.

Titman, Sheridan and Walter N. Torous. 1989. "Valuing Commercial Mortgages: An Empirical Investigation of the Contingent-Claims Approach to Pricing Risky Debt." *Journal of Finance* 44(2): 345-373.

Van Order, Robert and Peter Zorn. 2000. "Income, Location and Default: Some Implications for Community Lending." *Real Estate Economics* 28(3): 385-404.

Vandell, Kerry D. 1995. "How Ruthless Is Mortgage Default? A Review and Synthesis of the Evidence." *Journal of Housing Research* 6(2): 245-264.

Vandell, Kerry D. and Thomas Thibodeau. 1985. "Estimation of Mortgage Defaults Using Disaggregate Loan History Data." *Real Estate Economics* 13(3): 292-316.

Vandell, Kerry, Walter Barnes, David Hartzell, Dennis Kraft, and William Wendt. 1993. "Commercial Mortgage Defaults: Proportional Hazards Estimations Using Individual Loan Histories." *Real Estate Economics* 21(4): 451-480.

Vassalou, Maria and Yuhang Xing. 2004. "Default Risk in Equity Returns." *Journal of Finance* 59(2): 831–868.

White, Brent T. 2010b. "Underwater and Not Walking Away: Shame, Fear and the Social Management of the Housing Crisis." *Wake Forest Law Review* 45(4): 971-1023.

Wu, Jing, Joseph Gyourko, and Yongheng Deng. 2012. "Evaluating Conditions in Major Chinese Housing Markets." *Regional Science and Urban Economics* 42(3): 531-543.

Zmijewski, Mark E. 1984. "Methodological Issues Related to the Estimation of Financial Distress Prediction Models." *Journal of Accounting Research* 22: 59-82.